AN INVESTIGATION INTO THE MOLECULAR GENETICS OF THE INFLAMMATORY BOWEL DISEASES

COLIN LAMONT NOBLE

DECLARATION

I declare that the work contained within this thesis is original and has been composed by myself unless where clearly indicated. The research was undertaken in the Gastrointestinal Unit, University of Edinburgh, Western General Hospital, Edinburgh and at Genenetech Inc. South San Francisco, USA who funded the microarray study. I spent 20 months of the 2 years in Edinburgh and 4 months working in the pathology, microarray and bioinformatics departments at Genentech Inc. The data presented here has not been submitted for any other professional degree.

SUMMARY

The aims of this thesis were firstly to investigate gene expression profiles in human colonic and terminal ileal biopsies using microarray technology in a well phenotyped cohort of patients with Crohn's disease, ulcerative colitis and a control cohort. The role in disease pathogenesis of differentially expressed genes was investigated along with the expression of candidate genes identified by genome wide association study and cell lineage analysis. Parallel studies attempted to replicate the *Nature Genetics* publications of Peltekova and Stoll and colleagues who investigated the role the IBD5 locus and the DLG5 gene respectively in patients with inflammatory bowel disease.

In the healthy adult colon cluster analysis showed differences in gene expression between the right and left colon. (χ^2=25.1, p<0.0001). Developmental genes HOXA13, ($p=2.3x10^{-16}$), HOXB13 ($p<1x10^{-45}$), GLI1 ($p=4.0x10^{-24}$), and GLI3 ($p=2.1x10^{-28}$) primarily drove this separation.

Upregulated genes in the Crohn's disease biopsies compared to the controls included SAA1 (Fold change (FC) +7.5, $p= 1.47x10^{-41}$) and REGL (FC +7.3, $p =2.3x10^{-16}$). Cellular detoxification genes including-SLC14A2 (FC -2.49, p= 0.00002) were downregulated. In the Crohn's disease terminal ileal biopsies diubiquitin (FC+11.3, $p<1x10^{-45}$), MMP3 (FC +7.4, $p= 1.3x10^{-11}$) and IRTA1 (FC -11.4, $p= 4.7x10^{-12}$) were differentially expressed compared to controls. In the colon SAA1 (FC +6.3, $p= 5.3x10^{-8}$) was upregulated and TSLP (FC -2.3, $p= 2.7x10^{-6}$) was downregulated comparing non-inflamed Crohn's disease and control biopsies.

Of the Crohn's disease susceptibility genes identified by genome wide association scan IL-23A, JAK2 and STAT3 were upregulated in Crohn's disease confirming dysregulation of Th17 signalling. Modest differential expression was also observed in a number of the autophagy genes, notably ATG16L1. When clustering analysis was undertaken, terminal ileal Crohn's disease and terminal ileal control biopsies separated from colonic Crohn's disease and colonic control biopsies. Further clustering analysis of the terminal ileal biopsies showed separation between the terminal ileal Crohn's disease and control biopsies.

When the ulcerative colitis biopsies and control biopsies were compared, differentially upregulated genes in ulcerative colitis included SAA1 ($p<10^{-45}$) the alpha defensins, DEFA5&6 ($p=0.00003$ and $p=6.95x10^{-7}$ respectively), MMP3 ($p=5.6x10^{-10}$) and MMP7 ($p=2.3x10^{-7}$). Increased DEFA5&6 expression was further characterized to Paneth cell metaplasia by immunohistochemistry and *in-situ* hybridization.

Variants in all the examined IBD5 SNPs were associated with Crohn's disease ($p<0.003$). The IBD5 locus was also associated more severe Crohn's disease behaviour. In the absence of the IBD5 risk haplotype, no association of OCTN1/2 variants with Crohn's disease was detected. The analysis of the DLG5 variant 113A showed there were no associations with inflammatory bowel disease, Crohn's disease or ulcerative colitis.

Overall these data emphasise the key role of a number of inflammatory molecules and pathways in pathogenesis of Crohn's disease and ulcerative colitis, and their potential

for translation to therapeutic targets. The results also add considerably to the recent genome wide association studies in providing complimentary human colonic and ileal expression data along with detailed analysis of the IL-23 and autophagy pathways.

CONTENTS

	Page
Title	1
Declaration	2
Summary	3
Contents	6
Figures	13
Abbreviations	15
Publications	17
Abstracts	18
Oral Presentations	20
Acknowledgements	21
CHAPTER 1: INTRODUCTION	**23**
1.1 History of Crohn's disease and ulcerative colitis	24
1.2 Epidemiology of Crohn's disease and ulcerative colitis	26
1.3 Clinical presentation of inflammatory bowel disease	27
1.4 Evidence for the genetic basis of inflammatory bowel disease	28
1.5 Disease pathogenesis	29
1.6 MOLECULAR STRATEGIES FOR THE IDENTIFICATION OF SUSCEPTIBILITY GENES	**36**
1.6.1 Positional cloning genotyping	36
1.6.2 Positional candidate gene approach	37
1.6.3 Discovery of the NOD2/ CARD15 gene	37
1.7 THE IBD5 LOCUS	**39**
1.7.1 Function of OCTN1/2	42
1.7.2 Other candidate genes within the IBD5 locus	42

1.7.3 Genotype phenotype associations with IBD5- Age of diagnosis 43

1.7.4 Disease location 44

1.7.5 Disease behaviour 45

1.7.6 Other phenotypic associations 45

1.7.7 Epistasis between the IBD5 locus and NOD2/ CARD15 45

1.8 DROSOPHILA DISCS LARGE HOMOLOG 5 (DLG5) 47

1.8.1 Phenotypic associations with DLG5- gender stratification 51

1.8.2 Influence of DLG5 113A/R30Q stratified by social affluence and disease location 52

1.8.3 DLG5 epistasis with NOD2 /CARD15 53

1.8.4 Function of DLG5 54

1.9 GENOME WIDE ASSOCATION STUDIES 56

1.10 MICROARRAY GENE EXPRESSION STUDIES 60

1.10.1 History of micorarray technology 60

1.10.2 Basic microarray function 61

1.10.3 Development of microarray technology 63

1.10.4 Experiment design and interpretation 65

1.10.5 Validation of microarray results 66

1.10.6 Statistical analysis of microarray data 68

1.10.7 Gene ontology 69

1.10.8 Future implications 71

1.11 CLINICAL DATA GENERATED BY MICROARRAY STUDIES 72

1.11.1 Microarray studies in patients with inflammatory bowel disease 73

1.12 OTHER MOLECULAR IDENTIFICATION STRATIGIES 79

1.12.1 Proteomics 79

1.12.2 Yeast two- hybrid screening 80

CHAPTER 2: THESIS AIMS 82

CHAPTER 3: MATERIALS AND METHODS 85

3.1.1 Equipment 86

3.1.2 Reagents 87

3.2 MICROARRAY METHODS 89

3.2.1 Overview 89

3.2.2 Recruitment of patients for endoscopic biopsy collection 89

3.2.3 Microarray ethical approval 90

3.2.4 Biopsy collection 91

3.2.5 RNA isolation 91

3.2.6 Low input fluorescent linear amplification 92

3.2.7 Purification of amplified cRNA 95

3.2.8 Agilent oligonucleotide microarray hybridization 96

3.2.9 Stabilization and drying protocol 97

3.3.1 Real time PCR 97

3.3.2 In situ hybridization for defensin alpha 5 99

3.3.3 Immunohistochemistry for rabbit anti-human lysozyme and rabbit anti- human defensin alpha 6 100

3.3.4 Statistical analysis 101

3.4 GENOTYPING METHODS 103

3.4.1 Recruitment of patients and controls 103

3.4.2 Phenotypic data acquisition 103

3.4.3 Ethics approval 104

3.4.4 DNA extraction and storage 104

3.4.5 PCR reaction: TaqMan® technology 105

3.4.6 Allelic discrimination in TaqMan® 106

3.4.7 Genetic data analysis 107

CHAPTER 4: CHARACTERISATION OF INTESTINAL GENE EXPRESSION PROFILES IN CROHN'S DISEASE BY GENOME-WIDE MICROARRAY ANALYSIS **109**

4.1 INTRODUCTION **111**

4.2 MATERIALS AND METHODS **113**

4.2.1 Biopsy collection 113

4.2.2 Microarray analysis 115

4.2.3 Real Time PCR 115

4.2.4 Data analysis 116

4.3 RESULTS **117**

4.3.1 Gene expression in Crohn's disease and controls 118

4.3.2 Gene expression in the terminal ileum 120

4.3.3 Colonic gene expression analysis 123

4.3.4 Analysis of gene expression changes in inflammatory bowel disease biopsies and controls 125

4.3.5 Expression of genes implicated by GWAS meta-analysis 125

4.3.6 The IL-23 pathway 127

4.3.7 Autophagy pathway 128

4.3.8 Unsupervised hierarchical clustering analysis 124

4.3.9 Hierarchical clustering by specific probe subsets: *Immune Response in Silico (IRIS)* probes 134

4.3.10 Hierarchical clustering by epithelial cell markers 134

4.3.11 Real time PCR confirmation of microarray results 136

4.4 DISCUSSION **138**

CHAPTER 5: REGIONAL VARIATION IN GENE EXPRESSION IN THE HEALTHY COLON IS DYSREGULATED IN ULCERATIVE COLITIS **145**

5.1 INTRODUCTION **147**

5.2 METHODS **148**

5.2.1 Patients and controls 148

5.2.2 Biopsy collection 149

5.2.3 Microarray analysis 149

5.2.4 Real time PCR, I*n Situ* hybridization for defensin alpha 5 and immunohistochemistry for rabbit anti-human lysozyme and rabbit anti-human defensin alpha 6 150

5.2.5 Data analysis 150

5.3 RESULTS **151**

5.3.1 Influence of anatomical location on gene expression in the healthy colon and terminal ileum 151

5.3.2 Analysis of expression in ulcerative colitis and control biopsies 153

5.3.3 Analysis of expression in sigmoid colon biopsies in patients with quiescent ulcerative colitis and non-inflamed control biopsies 153

5.3.4 Inflamed versus non-inflamed ulcerative colitis sigmoid colon biopsies 158

5.3.5 Analysis of specific gene families- alpha defensins 5 and 6 158

5.3.6 ATP- binding cassette (ABC) transporter family and the Xenobiotic- transcription regulators 161

5.3.7 RTPCR analysis 161

5.3.8 *In-Situ* hybridization and immunohistochemistry 165

5.3.9 Expression of genes within the IBD2 locus 165

5.4 DISCUSSION **169**

CHAPTER 6: THE CONTRIBUTION OF OCTN1/2 VARIANTS WITHIN THE IBD5 LOCUS TO DISEASE SUSCEPTIBILITY AND SEVERITY IN CROHN'S DISEASE **175**

6.1 INTRODUCTION **177**

6.2 METHODS **178**

6.2.1 Patients and controls 179

6.2.2 Genotyping 179

6.2.3 Data analysis 181

6.2.4 Expression of genes within the IBD5 locus 182

6.3 RESULTS **183**

6.3.1 Linkage disequilibrium across the IBD5 locus 183

6.3 2 Analysis of SNPs representing the extended IBD5 haplotype (IGR2096, IGR2198 and IGR2230). 185

6.3.3 Analysis of OCTN1/2 variants and OCTN1/2 TC haplotype 188

6.3.4 The OCTN1/2 association with Crohn's disease is not independent of the association with the extended IBD5 haplotype. 190

6.3.5 Expression of genes within the IBD5 locus 192

6.3.6 Age of diagnosis 195

6.3.7 Anatomical distribution 195

6.3.8 Disease behaviour 196

6.3.9 Disease Progression 198

6.3.10 Requirement for surgery 198

6.4 DISCUSSION **200**

CHAPTER 7: DLG5 VARIANTS DO NOT INFLUENCE SUSCEPTIBILITY TO INFLAMMATORY BOWEL DISEASE IN THE SCOTTISH POPULATION **206**

7.1 INTRODUCTION **208**

7.2 PATIENTS AND METHODS **210**

7.2.1 Demographics: Crohn's disease and ulcerative colitis 210

7.2.2 Control subjects 210

7.2.3 Genotyping 212

7.2.4 Data analysis 213

7.2.5 Expression of the DLG5 probe A_23_P161209 213

7.3 RESULTS 214

7.3.1 Disease susceptibility: haplotype D (113A) 214

7.3.2 DLG5 Haplotype A 215

7.3.3 Phenotypic analysis 216

7.3.4 Expression of DLG5 216

7.4 DISCUSSION 218

CHAPTER 8: CONCLUSIONS AND FUTURE DIRECTIONS 223

8.1 Conclusions 224

8.2 Future Directions 228

9.0 REFERENCES 230

APPENDIX 1 SUPPLEMENTARY TABLES COMPACT DISC (CD)

APPENDIX 2 PUBLICATIONS

FIGURES

		Page
Chapter 1		
1.1	Schematic of gene expression analysis using oligonucleotide microarray technology.	62
1.2	A gene ontology map involving a subset of differentially regulated genes in endoscopic biopsies of patients with inflammatory bowel disease.	70
Chapter 3		
3.1	*Agilent* low input RNA amplification process	94
3.2	TaqMan® allelic discrimination plot	107
Chapter 4		
4.1	Expression analysis of the IL-23/ Th17 pathway in Crohn's disease and controls	128
4.2	Expression analysis of the autophagy pathway in Crohn's disease and controls	130
4.3	Hierarchical clustering of terminal ileal biopsies from females with Crohn's disease and controls	133
4.4	Sigmoid colon Crohn's disease and control biopsies clustered by epithelial cell markers	135
Chapter 5		
5.1	Unsupervised hierarchical clustering of the control biopsies showing the location that the biopsy was retrieved from.	152
5.2	Expression of defensins alpha 5 and 6 in ulcerative colitis patients and controls.	160
5.3	Real time PCR expression of SAA1, IL-8 defensin alpha 5 and defensin alpha 6 in control and ulcerative colitis inflamed and non-inflamed sigmoid colon biopsies.	163
5.4	Real time PCR expression of MMP3, MMP7, S100A8 andTLR4 in control and ulcerative colitis inflamed and non- inflamed sigmoid colon biopsies.	164
5.5	Defensin alpha 5 in-situ hybridization in the terminal ileum and colon of patients with ulcerative colitis and controls.	166

5.6 Defensin alpha 6 immunohistochemistry in the terminal ileum and colon of patients with ulcerative colitis and controls. 168

Chapter 6

6.1 Haplotype structure and linkage disequilibrium across the IBD5 region. 184

ABBREVIATIONS

Apaf-1: Apoptotic peptidase activating factor 1,

ATG16L1: Autophagy-related protein 16-1

Aza: Azathioprine

CD: Crohn's disease

DEFα: Alpha defensin

DLG5: Drosophila Discs large homolog

DNA: Deoxyribonucleic acid

FDR: False discovery rate

GLI: Glioma-associated oncogene homolog

GWAS: Genome wide association study

HOX: Homebox genes

IBD: Inflammatory bowel disease

IBD5: Inflammatory bowel disease 5 locus

IL: Interleukin

LLR Leucine rich region

LOD: Logarithm (base 10) of odds

JAK2: Janus kinase 2

MAQC: Microarray quality control project

MDR1 (ABCB1): ATP-binding cassette, sub-family B member 1

MMF: Mycophenolate mofetil

MMP: Matrix metalloproteinase

MTX: Methotrexate

NOD2/CARD15: Nucleotide binding oligomerization domain containing 2

OCTN: Organic cation transporter

PAR: Population attributable risk

PDZ: Post synaptic density protein, Drosophila disc large tumor suppressor, and zonula occludens-1 protein

PCR: Polymerase chain reaction

RNA: Ribonucleic acid

SAA1: Serum amyloid A1

SCCAI: Simple clinical colitis activity index

STAT3: Signal transducer and activator of transcription 3

TI: Terminal ileum

TLRs: Toll like receptors

TSLP: Thymic stromal lymphopoietin

UC: Ulcerative colitis

x g: Times gravity

6MP: 6- Mercaptopurine

PUBLICATIONS

Noble CL, Abbas AR, Lees CW, Cornelius J, Toy K, Modrusan Z, Clark H, Arnott ID, Penman ID, Satsangi J, Diehl L. Characterisation of Intestinal Gene Expression Profiles in Crohn's Disease by Genome- wide Microarray Analysis. Submitted to IBD.

Aldhous MC, **Noble CL**, Satsangi J. An investigation of colonic human β-Defensin 2 (HBD2) protein in inflammatory bowel disease. PLoS One. 2009 Jul 20;4(7):e6285.

Lees CW, Zacharias W, Tremelling M, **Noble CL**, Nimmo ER, Tenesa A, Cornelius J, Ho GT, Torkvist L, Linforss U, Lofberg R, Farrington S, Fitch P, Drummond H, Arnott ID, Diehl L, Campbell H, Dunlop MG, Parkes M, Howie SE, Gumucio DL, Satsangi J. The Hedgehog signalling pathway is dysregulated in ulcerative colitis. PLoS Med. 2008 Dec 9;5(12):e239.

Noble CL, Abbas A, Cornelius J, Lees CW, Ho GT ,Toy K, Modrusan Z, Pal N, Zhong F, Chalasani S, Clark H, Arnott ID, Penman ID, Satsangi J, Diehl L. The Characterisation of Distinct Intestinal Gene Expression Profiles in Ulcerative Colitis by Microarray Analysis. Gut. 2008 Oct;57(10):1398-405.

Browning BL, Barclay M, Bingham SA, Brand S, Büning C, Castro M, Drummond H, Ferguson L, Fisher S, Gearry R, Glas J, Henckaerts L, Huebner C, Lakatos L, Lakatos P, Latiano A, Liu X, Mathew C, Müller-Myhsok B, Newman W, Nimmo E, **Noble CL,** Parkes M, Petermann I, Rutgeerts P, Satsangi J, Shelling A, Siminovitch K, Török H, Tremelling M, Vermeire S, Vito V, Witt H, Gender-stratified analysis of DLG5 R30Q in 4707 Crohn's disease patients and 4973 controls from 12 Caucasian cohorts. J Med Genet. 2008 Jan;45(1):36-42.

Satsangi J, **Noble C**, Arnott ID, Russell R Nimmo ER, DLG5 variants and susceptibility to inflammatory bowel disease in the Scottish population, Authors' reply. Gut 2006 Jul; 55(7):1050.

Noble CL, Nimmo ER, Gaya D, Russell RK, Satsangi J. Novel Susceptibility Genes in Inflammatory Bowel Disease. World Journal of Gastroenterology 12(13):1991-99.

Tenesa A, **Noble C,** Satsangi J and Dunlop M. Association of DLG5 and inflammatory bowel disease across populations. E. J. Hum. Genent. 4-1-2006.

Noble CL, Nimmo ER, Drummond H, Ho GT, Tenesa A, Smith L, Anderson N, Arnott IDR, Satsangi J. Analysis of the contribution of OCTN1 / 2 variants within the IBD5 locus to disease susceptibility and severity in Crohn's disease. Gastroenterology 2005:129: 1845-1864.

Noble CL, Nimmo ER, Drummond H, Smith L, Arnott ID, Satsangi J.
DLG5 variants do not influence susceptibility to inflammatory bowel disease in the Scottish Population. Gut 2005:54:1416-1420.

ABSTRACTS

Aldhous MC, **Noble CL**, Satsangi J. The regulation of colonic human β-Defensin 2 (HBD2) production in inflammatory bowel disease. Gut, 2009, 58(1), A13.

Lees CW, **Noble CL**, Diehl L, Satsangi J. Expression analysis of all genes implicated in susceptibility to Crohn's disease on genome- wide association studies. Gastroenterology, 2008, 134(4), A41.

Lees CW, **Noble CL**, Diehl L, Satsangi J. Expression analysis of all genes implicated in susceptibility to Crohn's disease on genome- wide association studies. Gut, 2007, 57(1), A121.

Noble CL, Abbas A, Cornelius J, Penman I, Arnott IDR, Modrusan Z, Toy K, Diehl L, Satsangi J. Distinct intestinal gene expression profiles in the terminal ileum of patients with Crohn's disease. Gastroenterology, 2007: 132 (4) A447.

Ho GT, **Noble CL**, Cornelius J, Penman ID, Arnott ID, Diehl L, Satsangi J. The human ATP- binding cassette (ABC) transporters superfamily and xenobiotic-transcription regulators: Analysis of intestinal epithelial gene expression in inflammatory bowel disease. Gastroenterology, 2007: 132 (4) A8.

Noble CL, Abbas A, Cornelius J, Penman I, Arnott IDR, Modrusan Z, Toy K, Diehl L, Satsangi J. Analysis of distinct intestinal gene expression profiles in the terminal ileum of patients with Crohn's disease. Gut, 2007, 56(11), A116.

Ho GT, **Noble CL,** Cornelius J, Penman ID, Arnott ID, Diehl L, Satsangi J. The human ATP- binding cassette (ABC) transporters superfamily and xenobiotic-transcription regulators: Analysis of intestinal epithelial gene expression in inflammatory bowel disease. Gut, 2007, 56(11), A111.

Noble CL, Abbas A, Cornelius J, Penman I, Arnott IDR , Modrusan Z, Toy K, Diehl L, Satsangi J. The Characterisation of intestinal gene expression profiles in inflammatory bowel disease by microarray analysis. Gastroenterology, 2006: 130(4), A65.

Noble CL, Nimmo ER, Russell R, Drummond H, Smith L, Arnott IDR, Satsangi J. Analysis of the contribution of DLG5 in inflammatory bowel disease. Gastroenterology, 2005: 128(4), A446.

Noble CL, Nimmo ER, Russell R, Drummond H, Smith L, Arnott IDR, Satsangi J. The Contribution of Octn1/2 Variants within the IBD5 Locus To Disease Susceptibility and Severity in Crohn's Disease. Gastroenterology, 2005: 128(4), A446.

Noble CL, Nimmo ER, Drummond H, Smith L, Arnott IDR, Satsangi J. Analysis of the contribution of OCTN1/ 2 polymorphisms within the IBD5 to susceptibility and severity in Crohn's disease. Gut. 2005;54(11), A96.

Noble CL, Nimmo ER, Drummond H, Smith L, Arnott IDR, Satsangi J. OCTN 1& 2 Polymorphisms within the IBD5 locus predict susceptibility and severity in Crohn's Disease. Aliment.Pharmacol.Ther. 2005;21, 199.

ORAL PRESENTATIONS

Lees CW, **Noble CL**, Diehl L, Satsangi J. Expression analysis of all genes implicated in susceptibility to Crohn's disease on genome- wide association studies. Digestive Diseases Week, San Diego, 2008.

Noble CL, Abbas A, Cornelius J, Penman I , Arnott IDR , Modrusan Z, Toy K, Diehl L, Satsangi J. The Characterisation of intestinal gene expression profiles in inflammatory bowel disease by microarray analysis. Digestive Diseases Week, Los Angeles, 2006.

Noble CL, Nimmo ER, Drummond H, Smith L, Arnott I, Satsangi J. The IBD5 locus influences disease susceptibility and behaviour in Crohn's disease. Scottish Society of Gastroenterology. Glasgow. November, 2004.

ACKNOWLEDGEMENTS

I would like to thank my two supervisors Jack Satsangi and Lauri Diehl for all the help and guidance I have received. I spent 20 months of the 2 years at the Gastrointestinal Unit, Western General Hospital, Edinburgh, obtaining ethics, collecting biopsies and phenotypic data, and undertaking analysis of the microarray and genetic data sets. This was supervised by mainly by Jack Satsangi. The other 4 months of the 2 years were spent at Genenetech Inc, South San Francisco carrying out the microarray experiments and data analysis, and this was supervised by mainly by Lauri Diehl.

I would like to acknowledge Jennine Cornelius for carrying out and teaching me how to isolate RNA from the biopsies and the real time PCR. I would also like to acknowledge Karen Toy for carrying out and showing my how to amplify the RNA, how to purify it, and the hybridization, stabilization, drying and scanning steps. I would like to thank Navneet Pal and Fiona Zhong for carrying out the in Situ Hybridization for Defensin Alpha 5 and Sreedevi Chalasani for carrying out the immunohistochemistry for Defensin Alpha 6. Alex Abbas provided help with the data analysis.

For the genetics studies I would like to thank Ian Arnott, Jack Satsangi, Janice Fennell and Linda Smith for recruiting patients. I would also like to thank Ian Arnott, Hazel Drummond, Janice Fennell and Gwo-Tzer Ho for collecting phenotypic data. I would like to acknowledge Norman Anderson for extracting the DNA and Angela Fawkes and Stuart Bayliss for the genotyping. I would also like to thank Hazel Drummond for

maintaining the database. Finally, I would also like to thank my wife Katharine and my daughters Anna and Zoë for their gentle encouragement and support.

CHAPTER 1

INTRODUCTION

1.1 History of Crohn's disease and Ulcerative colitis

Although not labelled Crohn's disease, descriptions of regional intestinal inflammation have been documented for at least 200 years. Perhaps the most detailed early description of Crohn's disease was written in 1913 by Thomas Kennedy Dalziel, an Edinburgh trained surgeon who was working in Glasgow. He observed two patients who had died as a result of gross small bowel disease and 7 patients with diseased small bowel segments that were able to be removed surgically, resulting in a clinical recovery.(1)

In the USA in the 1920s and 1930s increasing numbers of young patients were being referred with symptoms of fevers, abdominal cramps, weight loss and diarrhoea. In 1923 at the Mt Sinai Hospital in New York three surgeons Berg, Oppenheimer and Ginzberg, collated a case series of 12 patients who presented with these symptoms and at surgery had small bowel inflammation that did not fit with any known disease.

These cases were combined with two further cases under the care of Dr Burril Crohn and Crohn, Ginzburg and Oppenheimer published their findings in the Journal of the American Medical Association in 1932.(2) They described a disease if the terminal ileum- 'terminal ileitis' as a clinical entity, however it was quickly realised that the disease could present in the jejunum and colon resulting in the name 'regional enteritis' being adopted. Over the years this has fallen from favour and the name 'Crohn's disease' has been adopted. The involvement of peri- anal disease was also recognised as a phenotype of Crohn's disease in 1938.(3)

The initial management of Crohn's disease was largely surgical or conservative with medical therapy relying on nutritional support, correction of anaemia, antibiotics and in the 1950s the administration of ACTH or corticosteroids.(4) Even with the modern armourmentarium of medical therapy the debate continues about the optimal management of Crohn's disease. At the forefront of this debate is the question of whether to persue a step up approach escalating treatment in response to symptoms or to try to alter the natural history of the disease by embarking on early aggressive medical therapy from diagnosis.(5)

Infectious and non- infectious bloody diarrhoea have existed and been documented for hundreds of years as a clinical syndrome. Ulcerative colitis was first described as a distinct clinical entity by Samuel Wilkes a physician at Guy's Hospital in London in 1859. In his initial letter to the editor of the *Medical Times and Gazette* he described the post-mortem appearances of a colon of a patient with ulcerative colitis. The disease was further characterised in a cohort of 300 patients from the London Teaching Hospitals in 1909 and it was observed that the disease ran a relapsing remitting course and that the first attacks were often the most severe.(6)

Prior to the 1950s the mainstay in the management of ulcerative colitis was surgery, however the introduction of corticosteroids in the early 1950s had a dramatic effect in lowering the mortality from exacerbations of ulcerative colitis from 37% to less than 1%.(7) Sulphasalazine was the first drug to be used in the aminosalicylate class that mediate their anti-inflammatory action via 5-aminosalicylic acid (5-ASA) in the colon. This family of drugs have continued to form the mainstay of maintenance therapy in ulcerative colitis.(8) The thiopurine analogues azathioprine and 6-

mercaptopurine were introduced in the early 1970s further reducing the need for surgical intervention in ulcerative colitis, however it is now becoming apparent that in selected cases of difficult to treat ulcerative colitis elective surgery reduces mortality.(9)

1.2 Epidemiology of Crohn's disease and Ulcerative colitis

In the last century, despite factoring in the increasing awareness, case definition and improving diagnostic techniques there has been an ever increasing incidence of inflammatory bowel disease in the United Kingdom.(10) Seven prospective and retrospective studies over a period of greater than 20 years in the adult Crohn's disease population have shown an increase in the incidence of Crohn's disease in Northern Europe and North America in the 1960s and 1970s, but since the 1980s the incidence has largely plateaued.(11) Incidence rates in the UK have been estimated to be between 5.9 and 11.1 cases per 100 000 of the population.(12) The most recent data from Scotland was published in 1992 and in North East Scotland and the Northern Islands the incidence of Crohn's disease was observed to be 11.7 cases per 100 000 population.(13)

In the Scottish paediatric population (under 16), the incidence of Crohn's disease has increased by approximately 3 times between the 1970s and the 1980s paralleling the increase in the adult populations following a lag period.(14) Further data from our unit have shown a North to South gradient with higher incidences in the North- 3.2 per 100 000 versus the South 2.1 per 100 000.(15) The incidence of paediatric Crohn's disease in Stockholm doubled during the 1990s.(16) However, data from

England showed no increase in incidence during the 1990s (17) and differences in these data may be due to an earlier diagnosis of Crohn's disease or changes in diagnostic criteria.

There have been fewer studies on the incidence of ulcerative colitis probably due to the spectrum of disease and the difficulty in distinguishing a single episode of ulcerative colitis from other aetiologies. That said there has been a modest increase in the incidence of ulcerative colitis in the adult population (11) and data from faecal occult blood testing for colon cancer in Nottingham in previously undiagnosed subjects put the prevalence of ulcerative colitis at 700 per 100 000 of the population.(18) The incidence of ulcerative colitis in the Scottish paediatric population has also been steadily increasing.(19) Overall these data suggest that the increasing incidence of inflammatory bowel disease is strongly indicative of a change in the environment that we live in and this allied to genetic susceptibility has led to the observed increases.

1.3 Clinical Presentation of Inflammatory Bowel Disease

Crohn's disease can affect any part of the GI tract from the mouth to the anus. Symptoms at presentation include diarrhoea, abdominal pain, weight loss and in the case of peri- anal Crohn's disease pain or discharge around the anus. Patients can also have extraintestinal manifestations including seronegative arthritis, uveitis and dermatological conditions such as erythema nodosum and pyoderma gangrenosum. In patients who present in adulthood with Crohn's disease the distribution of disease is largely in the terminal ileum, colon or a combination of the two.

In the paediatric population growth retardation is a common presenting complaint and in this population Crohn's disease is characterised by extensive intestinal involvement and rapid disease progression to more complex disease involving strictures and fistulae formation.(20) The inflammation involved in Crohn's disease is often transmural and discontinuous leading to complications such as strictures and fistulae formation whereas in ulcerative colitis the disease typically involves only the mucosal and sub-mucosal layers of the bowel.(21)

Ulcerative colitis affects the colon in a confluent manner from the rectum proximally and classically presents with diarrhoea mixed with blood and mucus. In adults around 1/3 of patients will have distal disease, 1/3 will have involvement of the left side of their colon and 1/3 will have extensive disease involving the majority of the colon. In the paediatric population ulcerative colitis involves the majority of the colon in 82% of the patients.(20)

1.4 Evidence for the genetic basis of inflammatory bowel disease

The most convincing evidence in support of a genetic contribution towards the development of inflammatory bowel disease is provided from twin concordance studies in the UK, Sweden and Denmark.(22-24) When data from these studies was combined the concordance rates in monozygotic twins with Crohn's disease was 19/53 (36%), compared to concordance rates of 4/99 (4%) in dizygotic twins with Crohn's disease.(25) In twins with ulcerative colitis concordance rates of 12/75 (16%) and 4/99 (4%) were observed in monozygotic and dizygotic twins respectively,(25) a weaker genetic effect than that which was observed in patients with Crohn's disease.

Moreover, when data from the Swedish(26) and Danish (22) cohorts were examined with regards to location and date of diagnosis in the Crohn's disease concordant pairs, 64% were concordant with regards to disease location and 57% had been diagnosed within two years of each other.(25)

Further studies examining multiply affected families would also suggest a strong genetic contribution towards the development of inflammatory bowel disease with sibling relative risks of 25-42 for Crohn's disease (27-30) and 8-15 for ulcerative colitis.(28;31) The twin of a patient with inflammatory bowel disease has the highest relative risk of developing inflammatory bowel disease, followed by a patient's sibling and overall the relative risk to a first degree relative (sibling, offspring or parent) of developing inflammatory bowel disease is 5-35 if the proband has Crohn's disease and 10-15 if the proband has ulcerative colitis.(25)

Whilst initial data supported the hypothesis that the segregation pattern observed in ulcerative colitis indicated that a major dominant gene had a role in the susceptibility to ulcerative colitis and that a major recessive gene had a role in susceptibility to Crohn's disease,(32) more recent data would suggest that the genetic contribution towards the development of these diseases is inherited in a non- mendelian polygenic fashion.(33)

1.5 Disease Pathogenesis

In the broadest terms the main hypothesis for the pathogenesis of inflammatory bowel disease is that in genetically susceptible individuals there is a persistent excessive

immune response to commensal gastrointestinal flora.(33;34) Within this hypothesis investigators continue to define the role of the gut microflora and the role of potential pathogenic bacteria along with the importance of gut barrier integrity in inflammatory bowel disease. At the barrier organ interface of the gastrointestinal tract where a single layer of epithelial cells, goblet cells, Paneth cells and dendritic cells separate the host from the intestinal milieu, tolerance is required to commensal bacteria along with the ability to neutralize pathogenic bacteria.

In order to maintain mucosal integrity in the gut continued sampling of the luminal contents is undertaken. This allows for the antigen detection which focuses the immune response and this process is dominated by the antigen presenting myeloid derived dendritic cells in the lamina propria.(35) Disruption of the epithelial barrier or the presence of pathogenic bacteria results in the activation of resident dendritic cells via a CX3CR1- dependent mechanism, the production of IL-23, intestinal inflammation and the migration of innate immune cells such as neutrophils, macrophages and dendritic cells.(36)

From an evolutionary perspective, the innate immune response is more ancient than the adaptive immune response and it is present in nearly all taxa.(37) As a further measure of the importance of the innate immune response in plants, genes involved in plant disease resistance occupy >1% of the genome of primitive species, such as *Rockcress.*(38)

Toll like receptors (TLRs) are transmembrane proteins which play a pivotal role in mediating the innate immune response to viral, bacterial and fungal pathogens.(39)

TLRs were initially identified in *Drosophila* and they have been widely conserved across animal species.(40) In mammals TLRs are integral membrane glycoproteins which recognise conserved products unique to microbial metabolism and signal via a number of downstream signalling molecules (MyD88, Il-1R-associated kinases, TGF-β, and TNF- receptor associated factor 6).(41) To date 11 members of the TLR family have been identified in mammals.(40) Initially it was thought that these receptors were only expressed on macrophages and dendritic cells but it is now recognized that they are present on a wide variety of cell types.(35)

The exciting finding of an association between variants in the NOD2/ CARD15 gene and Crohn's disease has also focused investigation on the NOD2 (nucleotide- binding oligomerization domain) – LLR (leucine rich region) protein family.(42;43) These proteins are key regulators of innate immunity and apoptosis in mammals and plants.(44) NOD-LLR proteins play a central role in recognizing pathogens at the cell surface and in the cytosol, and a family of greater that twenty human proteins that possess a NOD domain have been identified.(45) The NOD-LLR proteins are comprised of three distinct functional domains, the amino terminal effector domain which is involved in signalling, the centrally located NOD domain and the LLR ligand-sensing domain.(46) The effector domains are distinct linking these proteins to multiple signalling pathways. Three of the human NOD-LLR proteins have effector domains which are caspase-recruitment domains (CARD): NOD1/CARD4, NOD2/CARD15 and Apaf-1.(44)

NOD2/ CARD15 functions as an intracellular sensor of muramyl dipeptide, a highly conserved peptidoglycan motif common to many intraluminal bacteria.(47;48)

Watanabe and colleagues have suggested that NOD2/CARD15 -/- mice lose negative control of TLR2 mediated activation of NFκB, potentially offering an explanation for the Th1 phenotype characteristic of Crohn's disease.(49) However, more recently published studies do not provide support for NOD2/CARD15 interaction with the TLR2 pathway and collectively these data emphasise the complexity of NOD2/CARD15 activation.(50;51)

The discovery of ATG16L1 as a Crohn's disease specific susceptibility gene has strongly implicated the autophagy pathway in the pathogenesis of Crohn's disease. Autophagy is a highly conserved cellular process where the cell digests part of its own cytoplasm and it functions as a normal physiological response to remove toxic material or intracellular bacteria from the cell. The pathway has also been implicated in the pathogenesis of neurodegenerative diseases such as Alzheimers's and Parkinson's disease.(52)

Recent data have linked the innate immune response and the autophagy pathway via Toll-like-receptor (TLR) engagement.(53) TLR induced phagosomes within macrophages triggered ATG5 and ATG7 mediated acidification and enhanced killing of the ingested organisms. These interactions between the innate immune system and the autophagy pathway have provoked investigators to speculate about specific interaction between NOD2/ CARD15 and autophagy and this is an area of active investigation.

Following on from the activation of the innate immune response the adaptive immune response sustains the persistent inflammation and tissue damage observed in patients

with inflammatory bowel disease.(21) Previously it was thought that ulcerative colitis patients exhibited a Th2 T cell cytokine profile and Crohn's disease patients displayed a Th1 T cell cytokine profile, but it is now obvious that there is considerable overlap between the T cell responses in ulcerative colitis and Crohn's disease and that the Th17 T cell pathway is central to the inflammatory response observed in Crohn's disease.(35)

Th 17 T cells are derived from naive CD4 T cells and this differentiation is driven by the sequential action of transforming growth factor β, IL-6 and IL-23.(54) Production of IL-23 appears to be limited to dendritic cells and macrophages and this amplifies the IL-23/ Th 17 pathway at several levels.(55;56) Genetic and functional studies have now emphasised the importance of this pathway in the pathogenesis of inflammatory bowel disease (57;58) and therapeutic targets have been identified. Clinical trials of a monoclonal antibody against the p40 subunit of IL-23 have produced promising early clinical data.(59)

Elegant murine studies have demonstrated that in germ free environments genetically susceptible or inbred mice do not develop colitis, however following the introduction of specific bacteria or different combinations of bacteria, differing patterns of inflammation occur in the gut.(60) Increased virulence of commensal bacterial species such as *E Coli* can result in invasion of the epithelial layer thus stimulating a pathological response from the host which results in chronic inflammation and tissue injury. In humans it has been observed that invasive mucosal adherent bacteria are more common in Crohn's disease than in controls, but whether this is the primary insult or a secondary phenomenon remains unanswered.(61)

Impaired epithelial barrier function may also contribute to the pathogenesis of inflammatory bowel disease and enhanced mucosal permeability has been observed in a number of colitis susceptible mice models.(62) The tight junctions between the epithelial cells are dynamic and permeability is increased by pro- inflammatory mediators including TNF and IL-17.(35) In mucosal biopsies taken from patients with IBD it was observed that the junctional proteins E- cadherin and β catenin were downregulated .(62)

Specific interest has also focused on the role of the Claudin gene family in maintaining mucosal integrity.(62) Claudin 2 downregulates and redistributes tight junction components and Claudins 5 and 6 increase apoptosis of epithelial cells in Crohn's disease.(63) The importance of barrier exclusion thus reducing and neutralising translocated luminal bacteria has been observed in a number of mouse models and specific interest has focused on the multidrug resistance-1 gene (Mdr-1). Mdr-1 encodes p-glycoprotein which functions to clear microbial xenotoxins and Mdr-1 deletion in mice induces colitis.(64) Moreover, genetic variants in the Mdr-1 gene have been associated with ulcerative colitis.(65)

Integral to epithelial integrity and pathogen killing is the secretion of anti- microbial peptides by epithelial cells and Paneth cells. Levels of the human alpha defensins 5 and 6 have been observed to be reduced in patients with ileal Crohn's disease regardless of the degree of mucosal inflammation.(66) Furthermore in Crohn's disease patients who carry the NOD2/ CARD15 frame shift mutation Leu 1007 there is a further decrease in ileal defensin alpha 5 suggesting that NOD2/CARD15 function may affect alpha defensin secretion.(67)

More recently Simms and colleagues also showed that expression of DEFA5&6 was down regulated in terminal ileal CD biopsies.(68) However, this downregulation was inflammation specific probably reflecting a loss of the epithelial layer and a reduction of epithelial and Paneth cells as a consequence of persistent inflammation.

Taken together these data strongly support the hypothesis that inflammatory bowel disease results from a genetic predisposition and a consequent excessive immune response to normal gut flora and the challenge is to decipher the complex interplay between the host immune system and the gut microflora.

1.6 MOLECULAR STRATEGIES FOR THE IDENTIFICATION OF SUSCEPTIBILITY GENES

1.6.1 Positional Cloning Genotyping

Genome-wide linkage analysis using highly polymorphic microsatellite markers identified during the course of the Human Genome Project (69) has led to success in identifying genetic determinants in both single gene disorders and in complex genetic diseases. Successful genome scans have typically involved several hundred microsatellite markers and large numbers of multiply affected inflammatory bowel disease families (typically sibling pairs). The classical positional cloning approach uses this technique to examine whether the degree of sharing of variant alleles between affected individuals exceeds that which would be expected by chance alone.(70) These confirmed regions of linkage need to be further narrowed by fine mapping of these areas, as each spans a large genomic region.

The most widely accepted guidelines for assessing the results are those defined in 1995 by Lander and Kruglyak, who proposed criteria for reporting areas of linkage, (71) with areas of 'suggestive linkage' having LOD scores of 2.2 and above and p values of less than 7×10^{-4}, areas of 'significant linkage' having LOD scores above 3.6 and p values of less than 2×10^{-5}, and areas of 'highly significant linkage' having LOD scores of 5.4 and above and p values of less than 3×10^{-7}. Areas of 'confirmed linkage' were defined as areas of significant linkage that have been replicated in an independent cohort, with a nominal p value of less than 0.01. Prior to genome wide association studies loci with confirmed linkage were identified on chromosomes

1(72), 3 (IBD9)(73;74), 5 (IBD5)(75-77), 6 (IBD3; HLA)(76;78), 12 (IBD2)(79), 14 (IBD4)(75;80), 16 (IBD1)(79;81) and 19 (IBD6)(76).

1.6.2 Positional Candidate Gene Approach

The positional candidate gene approach to identifying association involves the use of data concerning gene function and expression allied to chromosomal location. In inflammatory bowel disease genes involved in the regulation of the immune system, mucosal integrity and cell-cell interactions are all clearly plausible candidate genes. By comparing the allelic frequencies of variants in these genes between patients with IBD and matched controls (case control analysis), or by investigating intra-familial association/ linkage, genes that contain critical disease causing mutations may be identified.

1.6.3 Discovery of the NOD2/ CARD15 Gene

The discovery of the NOD2/ CARD15 gene as the susceptibility gene in IBD1 serves as proof of principle of both the positional cloning and positional candidate techniques. The IBD1 locus located on chromosome 16 was first identified by two, two stage genome scans of inflammatory bowel patients in 1996.(79;81) This led to the discovery published in consecutive *Nature* publications in 2001 using a positional cloning and a positional candidate technique, that variants of the NOD2/ CARD15 gene located within IBD1 were associated with susceptibility to Crohn's disease.(42;43) Since the identification of the three single nucleotide polymorphisms (SNPs) that have been associated with susceptibility to Crohn's disease- Gly908Arg

and Arg702Trp and the frameshift mutation Leu1007fsinsC, extensive studies examining the frequencies of these variants in cohorts of patients with inflammatory bowel disease have been undertaken.

Reported carriage of the three major risk alleles of NOD2/CARD15 in Crohn's disease vary between 0 and 50.0%, with highest rates seen in central European populations, (42;82) while mutations are absent in Japanese and Chinese series.(83;84) In the Northern European Crohn's disease population- Sweden, Finland, Denmark and Scotland lower NOD2/CARD15 frequencies have been reported suggesting genetic heterogeneity within Europe.(85;86) The calculated population attributable risk, the difference in rates of a condition between an exposed and unexposed population, of these variants has been lower, 11.4% in Stockholm county (86) and 11% in Scotland (85) compared to those in central Europe (27%-33%).(42;87)

A meta-analysis of 42 studies in patients with Crohn's disease revealed that heterozygote carriage of one of the 3 major risk alleles of NOD2/CARD15 conferred a 2.39 fold increased risk of developing Crohn's disease and homozygous/ compound heterozygous carriage conferred a 17 fold increased risk of developing Crohn's disease. (88)

1.7 THE IBD5 LOCUS

In 1999 a genome wide scan was carried out in 222 individuals from 46 families (20 Jewish and 26 non-Jewish), together with a total of 65 sibling pairs diagnosed with Crohn's disease.(75) One of the suggested areas of linkage in the Jewish families was at chromosome 5q33-q35 (Maximum LOD score = 2.2, p = 0.0003). In a further genome wide scan carried out by Rioux and colleagues an area of linkage was observed at chromosome 5q31-33 and when the authors carried out high density mapping of this area a significant association was observed with Crohn's disease (LOD score 3.9).(76) The area of observed association spanned a large 18cM area of chromosome 5q31 and this area of association was labelled IBD5.

To better delineate this region, Rioux and colleagues genotyped 256 Canadian father-mother, child trios where the child had Crohn's disease and at least one of the parents was unaffected using 56 microsatellite polymorphisms with an average spacing of 0.35cM. (77) Tight linkage disequilibrium was observed across the area examined and two adjacent markers demonstrated significant overtransmission IRF1p1 (p = 0.00016) and D5S1984 (p = 0.00039). A 435 kb risk haplotype block was identified spanning a cytokine cluster that contained a number of attractive candidate genes, however, no candidate risk alleles were identified in these genes. In order to look for novel SNPs in the risk locus DNA was sequenced in eight individuals across the region. 651 SNPs were identified and genotyped in the Crohn's disease population and 11 of these SNPs identified a 250kb risk haplotype.(77)

This 250kb region was of particular interest as it contained a number of immunoregulatory genes including interferon regulatory factor 1, IL-3, IL-4, IL-5 and IL-13. High resolution haplotypic analysis of the 5q31 region using 103 common SNPs (minor allele frequency >5%) revealed 11 discrete haplotype blocks that measured tens of kilobases in length that had limited genetic diversity and were punctuated by apparent sites of recombination.(89)

The association between the IBD5 locus and Crohn's disease was confirmed in a German cohort of trios by transmission disequilibrium testing (TDT) ($p = 0.007$) and interestingly an association was also observed with ulcerative colitis ($p = 0.002$).(90) Further data confirming the association between IBD5 and Crohn's disease has also been shown in a European case control study, (91) a British family based association study, (92) and a British case control study.(93)

Interestingly, data from Japan have shown that IBD5 variants are rare in the Japanese population (<1% for Crohn's disease, ulcerative colitis and healthy controls),(94) suggesting IBD5 is not a major component of inflammatory bowel disease susceptibility in Japan. However, given the ethnic diversity and large number of patients involved in studies showing a positive association between Crohn's disease and IBD5, the IBD5 locus remains one of the most robust susceptibility loci associated with immune mediated complex disease genetics.

In 2004 Peltekova et al suggested that the organic cation transporters SLC22A4 (OCTN1) and SLC22A5 (OCTN2) harboured the critical mutations that were associated with disease susceptibility in the 250 kb IBD5 locus.(95) They constructed

a two allele **TC** haplotype (SLC22A4 exon 9 1672C→**T** and SLC22A5 promoter -207G→**C**) and went on to demonstrate in a cohort of Canadian Crohn's disease patients, that in patients who lacked the extended IBD5 risk haplotype- homozygous with regards to the IBD5 marker SNP IGR2078, the **TC** haplotype was more prevalent in Crohn's disease patients than in controls. Population attributable risk (PAR %) for the TC haplotype were estimated at 19% for heterozygotes and 27% for homozygotes and no individual allelic data were given.

Case control data from cohorts of inflammatory bowel disease patients from Scotland and Germany replicated the association between the IBD5 loci, OCTN1/2 and Crohn's disease, however, an independent effect of the OCTN1/2 variants was not observed.(96;97) An overall association between the IBD5 locus and susceptibility to Crohn's disease and ulcerative colitis was also noted in more recent data from England and in a cohort of patients from Scotland who were diagnosed with IBD under the age of 16 years.(98;99) However, again no independent association between the variants in the OCTN1/2 genes and Crohn's disease was observed.

Using applied regression-based haplotypic analysis in 1200 European case control pairs Fisher and colleagues observed a highly significant association between the IBD5 locus and Crohn's disease.(100) The addition of the OCTN1/2 variants to a null model that included the background risk haplotype did not significantly improve the model fit and the authors concluded that the molecular basis for Crohn's disease susceptibility at the IBD5 locus remains to be defined. Taken together these data suggest that it may be premature to confirm that the OCTN1 (SLC22A4 exon 9 1672C→T) and OCTN2 (SLC22A5 promoter -207G→C) polymorphisms are the

critical mutations that confer disease susceptibility within the IBD5 locus and that they may simply be markers for the risk haplotype.

1.7.1 Function of OCTN1/2

OCTN1 is a 551 amino acid protein that is strongly expressed in kidney, trachea, bone marrow and to a lesser extent the small bowel and it has been characterised as a carnitine transporter.(101) An intronic SNP (rs2268277) in the RUNX1 (Runt-related transcription factor 1) binding site of OCTN1 has been associated with susceptibility to rheumatoid arthritis in the Japanese population.(102) OCTN2 is a 557 amino acid protein that is 75.8% homogenous to OCTN1, and functional studies have shown it to be a high affinity sodium carnitine transporter, expressed in kidney, smooth muscle and heart tissue.(103) Peltekova and colleagues suggested that the OCTN1 variant (1672C→T) has altered function in fibroblasts in vitro, with variant forms having less affinity for carnitine and a greater affinity for tetraethyl ammonium and some xenobiotics, and that the OCTN2 variant (-207G→C) disrupted a heat-shock transcription factor binding site in fibroblasts in vitro.(95) No data demonstrating the altered function or expression of OCTN1/2 gene products in Crohn's disease have yet been provided.

1.7.2 Other Candidate Genes within the IBD5 locus

Other attractive candidate genes that are located within the IBD5 locus in close proximity to OCTN1/2 include interferon regulatory factor 1 (IRF1), PDLIM4 and CSF2. IRF1 is a transcriptional factor that is involved in the regulation of a number of

genes related to the innate and adaptive immune response.(104) It has been observed that IRF1 is up regulated in intestinal mucosal mononuclear cells in Crohn's disease patients.(105)

PDLIM4 also known as RIL is a member of the ALP (actinin-associated LIM protein) subfamily of proteins who share an N-terminal PDZ domain and a C-terminal LIM domain.(106) PDZ domains are protein-protein interaction modules that are crucial for the assembly of structural and signaling complexes and interestingly it has been suggested that polymorphisms in another PDZ containing protein DLG5 which will be discussed in more detail later in the introduction are associated with the development of Crohn's disease.(107;108)

CSF2 encodes for granulocyte macrophage-colony stimulating factor (GMCSF) a myeloid growth factor that plays a crucial role in the potentiation and phagocytic ability of circulating neutrophils and macrophages.(109) Clinical data have demonstrated that recombinant GMCSF improved disease severity and quality of life in patients with moderate to severe Crohn's disease,(110) and more recently it was shown that GMCSF was more effective than placebo for inducing corticosteroid free remission in patients with corticosteroid dependant Crohn's disease.(111)

1.7.3 Genotype Phenotype Associations with IBD5: Age of Diagnosis

Data from the original genome wide scan carried out by Rioux and colleagues showed that the strongest association between the IBD5 locus and Crohn's disease was observed in families where the sibling was diagnosed with Crohn's disease under the

age of 16.(76) Mirza and colleagues also observed a significantly earlier onset of Crohn's disease in patients who were homozygote for the IBD5 risk haplotype ($p = 0.0019$),(91) however, when all the studies are considered no consistent association between IBD5 variants and an earlier age of diagnosis has been observed. (96;97;99;112;113)

1.7.4 Disease Location

Phenotypic data looking for associations between disease location and IBD5 variants has been inconsistent. Associations between individual IBD5 alleles or the IBD5 haplotype and gastric antral,(98) ileal, (112) colonic,(97) and perianal (93) Crohn's disease have all been reported and this variation may reflect different phenotypic classification categories or the fact that IBD5 variants may be associated with multiple disease locations. Further standardisation of Crohn's disease location may help to resolve this question.(114)

1.7.5 Disease Behaviour

Probably the most robust phenotypic associations have been observed between IBD5 variants and disease behaviour. In the Scottish Crohn's disease population variant alleles have been associated with markers of disease severity, namely disease progression from Vienna classification (115) inflammatory disease behaviour at diagnosis to stricturing and penetrating disease behaviour at follow up the adult population.(96) In Scottish Crohn's disease patients diagnosed at aged 16 or under an

association between low weight, low height, low body mass index at diagnosis and IBD5 variants was observed.(98)

In an Italian Crohn's disease cohort, a significant association between the OCTN1/2 TC haplotype and the presence of perianal fistulae and stricturing-fistulizing disease behaviour was observed (116) and in a further cohort of Crohn's disease patients from Belgium an association between IBD5 variants and penetrating Crohn's disease was observed.(113) These consistent data would suggest that IBD5 variants are associated with a more severe Crohn's disease phenotype.

1.7.6 Other Phenotypic Associations

When response to infliximab was examined in a Cohort of Spanish patients with Crohn's disease, the frequency of the IBD5 homozygous mutant genotype was significantly increased in Crohn's disease patients lacking response to infliximab (RR = 3.88).(117) Consistent with data suggesting an association between IBD5 variants and a more severe disease behaviour, Palmieri and colleagues also observed an association between IBD5 variants and patients on immunosuppressive medications.(118)

1.7.7 Epistasis between the IBD5 locus and NOD2/ CARD15

A number of publications have investigated potential gene-gene interactions between the IBD5 locus and NOD2/ CARD15 in patients with Crohn's disease, with the majority finding no association.(90;93;96;98;99) However, in two data sets an

interaction between the IBD5 locus and NOD2/ CARD15 was observed in patients with ulcerative colitis and this interaction appeared to be specific to the Leu1007fsinsC variant of NOD2/ CARD15.(90;119)

1.8 DROSOPHILA DISCS LARGE HOMOLOG 5 (DLG5)

A locus on chromosome 10 was first implicated in genome wide scanning of 282 affected sibling pairs in a European cohort (United Kingdom, Germany and the Netherlands) in 1999 as conferring susceptibility to inflammatory bowel disease (LOD score 2.30). (78) By fine mapping this region in a German cohort on four occasions progressively narrowing the area of association and better delineating it with more microsatellite markers, Stoll and colleagues identified one marker- rs1344966 that was significantly associated with inflammatory bowel disease ($p = 0.0006$) and Crohn's disease ($p = 0.002$).(107) This marker was adjacent to the gene DLG5 and following the sequencing of exons 2-32 of the gene, 33 SNPs were identified within this gene. These SNPs were then genotyped in 457 trios with inflammatory bowel disease and 18 of these variant SNPs were significantly overtransmitted compared to the wild type SNPs in the patients with inflammatory bowel disease.

Tight linkage disequilibrium (LD) was observed across the DLG5 gene that fell significantly at the boundaries of the gene and Stoll and colleagues identified two extended DLG5 haplotypes that influenced disease susceptibility in the German population.(107) The first haplotype (named haplotype D) was especially notable for the presence of a G→A substitution at nucleotide 113 that resulted in an amino change at position 30 from Arginine to Glutamine (R30Q). On analysis of carrier frequency, Stoll found the 113A variant to be associated with Crohn's disease (25% Crohn's disease v 17% healthy controls, P=0.001) in a case control study and trends between 113A transmission and inflammatory bowel disease (P=0.09) and Crohn's

disease (p = 0.065) were observed on transmission disequilibrium testing. The second haplotype, (haplotype A) was tagged by eight marker SNPs and was observed to be significantly under-transmitted in the inflammatory bowel disease group (P=0.006) suggesting the haplotype may be protective.(107)

There have now been a number of studies in genetically diverse populations attempting to replicate these associations (Table 1.1). Daly and colleagues investigated the 113A/ R30Q haplotype and haplotype A in cohorts of patients with Crohn's disease from Canada, Italy and the United Kingdom.(120) The association between the 113A/ R30Q variant and Crohn's disease was replicated in one of the case control studies involving patients from Italy and Canada and in a family based study involving families from Canada and the UK, however, no association was observed in the other case control study of UK patients. Haplotype A was not found to confer protection and the authors postulated DLG5 constituted a true IBD risk factor of modest effect. Further analysis of these data would argue that there was significant heterogeneity between the control groups from these ethnically diverse populations and that pooling data may mask true changes among different populations.(121)

A number of other studies have failed to replicate the initial association between the 113A /R30Q variant of DLG5 and Crohn's disease. Torok and colleagues found no association between DLG5 variants and inflammatory bowel disease, Crohn's disease or ulcerative colitis in a case control study in Germany and this data was replicated in Scotland where again no associations were observed.(97;122) No association was also observed in two cohorts of inflammatory bowel disease patients from Belgium- 373

patients in transmission disequilibrium testing analysis and 608 patients and 305 controls in a case control analysis,(113) and more recent data from cohorts of inflammatory bowel disease patients from Canada,(123) Norway,(124) Germany and Hungary,(125) Hungary alone (126) and Britain (127;128) also showed no association between DLG5 variants and IBD5.

Table 1.1: Allelic Frequencies of the DLG5 113A /R30Q variant in case control studies from genetically diverse populations.

Cohort Examined	Control Allelic Frequency (%)	IBD Allelic Frequency (%) (p value)	Crohn's Disease Allelic Frequency (%) (p value)	UC Allelic Frequency (%) (p value)
Stoll et al (107)	9.0	13.2 (0.0023)		
Daly et al (120)	5.9 9.7	11.0 (0.003) 9.3 (0.66)		
Noble et al (122)	13.2	11.4 (0.30)	11.4 (0.37)	11.4 (0.34)
Torok et al (97)	10.7	6.9 (< 0.001)	5.7 (< 0.001)	10.6 (0.91)
Vermiere et al (113)	10.5	11.8 (0.40)	11.5 (0.54)	13.6 (0.20)
Newman et al (123)	8.7	8.2 (0.62)	7.8 (0.79)	8.9 (0.65)
Bunning et al (125)	9.8 11.3	7.4 (0.076) 8.2 (0.36)	6.8 (0.056) 7.6 (0.11)	9.1 (0.72) 8.9 (0.33)
Tremelling et al (127)	11.6	10.1 (0.42)	10.4 (0.49)	9.7 (0.37)
Lakatos et al (129)	15.0	11.8 (0.12)	11.2 (0.52)	14.6 (0.79)
Pearce et al (128)	10.1	9.9 (0.84)	9.7 (0.73)	10.2 (0.90)

A significant association was observed between the 113A /R30Q variant and inflammatory bowel disease by Stoll and colleagues and in one of the two case control studies carried out by Daly and colleagues. This association was not observed in other populations presented and interestingly in the Torok study a significant negative correlation was observed.

Interestingly no 113A /R30Q variants were found in the Japanese Crohn's disease population (130) and this finding was replicated in a cohort of inflammatory bowel disease patients from Greece.(131) These studies confirm data examining variants in NOD2 /CARD15,(86) SLC22A4 /SLC22A5,(132) and CCL20 (133) where in ethnically diverse inflammatory disease populations, marked differences in the frequency of variants in these genes exist and further illustrate the difficulty in pooling ethnically diverse populations.

No studies have replicated the protective effect of haplotype A demonstrated by Stoll and colleagues,(107) and interestingly in analysis of cohorts of patients with inflammatory bowel disease and Crohn's disease from Canada and Scotland there was a trend towards haplotype A being overrepresented in patients compared to controls.(122;123)

1.8.1 Phenotypic Associations with DLG5- Gender Stratification

Perhaps the most interesting genetic data to emerge about DLG5 since it was proposed as a susceptibility gene by Stoll and colleagues has come from analysis of variant carriage in male and female patients with inflammatory bowel disease. In a cohort of patients diagnosed with inflammatory bowel disease at the age of 16 or under from Scotland, Russell and colleagues observed an overtransmission of the 113A /R30Q allele in male patients with Crohn's disease compared to females- 29.3% vs. 16.9%, p=0.04 OR 2.04.(134)

These data were supported by further analysis of the original German cohort of patients used by Stoll and colleagues,(107) combined with the Italian and Canadian patients used by Daly

and colleagues.(120;135) A significant association was observed between the 113A /R30Q allele and males with Crohn's disease- $p < 0.001$, OR = 2.49, but not females with Crohn's disease- $p = 0.979$, OR = 1.01.(135) The association between the 113A /R30Q variant and Crohn's disease in male patients appeared to be primarily driven by the difference between the allelic frequencies of the 113A /R30Q DLG5 variant in controls-males 5.2%, females 11.3%, rather than Crohns disease patients-males 10.1%, females 10.9% and this was supported by data from a newborn non-inflammatory bowel disease screening study- newborn males 7.1%; newborn females 11%, $p = 0.036$.

A further meta- analysis of 12 case-control cohorts containing 4707 Crohn's disease patients and 4937 controls showed no male-female allelic frequency differences in the control populations and a trend towards the 113A/R30Q variant being less common in the female Crohn's disease patients.(136) When 5 further case- control cohorts were added to the existing data the 113A/R30Q variant was associated with a small reduction in risk in female Crohn's disease patients- OR = 0.86 (CI 0.76- 0.97).

1.8.2 Influence of DLG5 113A/R30Q Stratified by Social Affluence and Disease Location

Multifactorial analysis by Russell and colleagues suggested that in the Scottish paediatric Crohn's disease population, higher social class (increased affluence) was independently associated with carriage of variants of 113A /R30Q - $p = 0.001$, OR=6.92.(134) A borderline association between the 113A /R30Q DLG5 allele and inflammatory bowel disease overall which became more significant in patients with increasing affluence in the paediatric Scottish

population, together with the lack of association in the Scottish adult inflammatory bowel disease population may suggest a specific DLG5 113A /R30Q- environment interaction.(122;134)

Further associations have been observed between DLG5 113A /R30Q variants and ileal disease location, smoking behaviour (128) and steroid resistance,(129) however, these data have not been consistent and in the majority of studies examining disease phenotype no associations have been observed.(97;122;123;125)

1.8.3 DLG5 Epistasis with NOD2 /CARD15

Data investigating potential gene-gene interaction between DLG5 and NOD2 /CARD15 have been varied and this lack of consistency probably relates to the size of studies required to prove an epistatic effect, and the difficulty in deciphering genetic, biological and statistical epistasis.(137) In the initial study by Stoll and colleagues, Crohn's disease patients who carried one or more of the common NOD2/ CARD15 variants also had a significant increase in transmission of the DLG5 113A /R30Q allele compared to Crohn's disease patients who were wild type for all 3 NOD2 /CARD15 variants.(107) This finding was replicated by Vermiere and colleagues who observed a higher frequency of DLG5 113A /R30Q in patients with one or more variant in NOD2/ CARD15- 12.2% compared those with no NOD2 /CARD15 variants- 8.7%, $p = 0.033$.(113)

However, in 3 large well designed case control studies from Britain including 2800 patients with inflammatory bowel disease, no evidence of epistasis between DLG5 and NOD2 /CARD15 was observed (122;127;128) and these findings were replicated by 2 central European studies involving over 1400 patients with inflammatory bowel disease.(97;129) Newman and colleagues also found no evidence of epistasis between DLG5 and NOD2 /CARD15 in a Canadian cohort of patients with Crohn's disease.(123)

1.8.4 Function of DLG5

DLG5 is a member of the MAGUK (membrane associated gyanylate kinase) gene family which encode cell scaffolding proteins and are also involved in intracellular signal transduction.(138) MAGUKs interact with other proteins to create an assembly of large multi-protein complexes that bind transmembrane proteins at the cytoplasmic side to other signal transduction proteins, therefore creating a platform for specific signal interactions.(139;140) Thus, DLG5 is important in maintaining epithelial integrity and variants in the gene may encode proteins that alter epithelial integrity and affect disease pathogenesis.

The DLG5 gene spans 79KB and contains 32 exons which encode a 180kb protein containing 1900 amino acids.(141) DLG5 is expressed most strongly in placental tissue and less so in heart, skeletal muscle, liver, small bowel and colon.(141) Stoll and colleagues carried out *in silico* analysis that suggested that the 113A /R30Q variant probably impairs DLG5 scaffolding function.(107)

Recent data have shown expression of DLG5 in human colonic tissue and close homology of DLG5 to CARD10, CARD11 and CARD14.(142) It was observed that DLG5 contains a N-terminal caspase recruitment domain (CARD) domain, and the authors suggested it may function in the regulation of NFkB activation or caspase activation as part of host defence mechanism.

1.9 GENOME WIDE ASSOCATION STUDIES

By using data generated from the Human Genome and International HapMap projects investigators are now able to use genome wide chips containing 300 000 to 500 000 single nucleotide polymorphisms (SNPs) to interrogate large case control data sets.(69) At present European ancestry population genome wide scans have been estimated to cover 80% of the Human genome variation where the calculated linkage between the SNPs-linkage disequilibrium r^2 is greater than 0.8.(143) With these dense genotyping chips becoming increasingly available and cheaper to purchase well powered studies to investigate complex inherited diseases are now possible.

In Crohn's disease the first genome wide association study was carried out in the Japanese Crohn's disease population and in a 2 stage process initially involving 94 Crohn's disease patients and 80 000 SNPs and secondly 484 Crohn's disease patients, a variant in the tumour necrosis superfamily gene TNFSF15 was significantly associated with Crohn's disease.(144) This finding was replicated in 2 European populations significantly strengthening this association.(57)

In a US wide study of 567 non-Jewish, European ancestory Crohn's disease patients with a bias towards ileal Crohn's disease, genome wide scanning was carried out using a chip that contained 308 332 SNPs.(145) Following a Bonferroni correction for multiple comparisons 3 SNPs remained significant- 2 SNPs representing the NOD/CARD15 gene and a further nonsynonymous SNP (Arg→Gln) in the IL-23R gene on chromosome 1p31. In addition 9

other SNPs in this region were found to be in linkage with the index SNP and were associated with Crohn's disease. This finding was replicated in a panel of patients of Jewish ancestory using a case control method and in a mixed Jewish and non- Jewish family trio cohort using family based association testing.(145)

The association between IL-23R and Crohn's disease has now been replicated in a UK wide case- control panel of Crohn's disease patients using 8 SNPs to represent the IL-23 gene.(146) Further to this a lesser association was observed between the IL-23R variants and ulcerative colitis. Three subsequent genome wide association studies in Flemish, UK and Quebec Crohn's disease cohorts have also replicated the association between variants in the IL-23R gene and Crohn's disease.(57;147;148) Taken together these data emphasise the importance of the proinflammatory IL- 23 pathway in the pathogenesis of inflammatory bowel disease and the potential for therapeutic intervention.

The largest and most impressive genome wide scan to date was carried out in the UK using DNA from 14000 patients and 3000 controls encompassing 7 major diseases- bipolar disease, coronary artery disease, Crohn's disease, rheumatoid arthritis, type 1 and type 2 diabetes.(57) Of the 24 SNPs that were found to have an association with their respective disease with p value of less than $5x10^{-7}$, 9 were observed in Crohn's disease. The strongest association in Crohn's disease mapped to the gene ATG16L1, a gene that is part of the autophagosome pathway implicated in processing of intracellular bacteria. The previous associations with NOD2/ CARD15, IL- 23R, the IBD5 locus and the gene desert on chromosome 5p13.1 were replicated and novel associations were observed with SNPs representing the genes IRGM, a

gene also involved in autophagy, MST1, the protein of MST1 influences phagocytic activity of peritoneal macrophages, NKX2-3, part of the NKX family of transcription factors and PTPN2, a negative regulator of inflammatory response.

A 'second tier' of areas of linkage was also reported based on statistical significance and biological plausibility. Among the loci reported were the HLA region on chromosome 6, the candidate gene TNFAIP3 whose protein product inhibits NFκB dependent TNFα production and TNFSF15 where a previous association had been observed in the Japanese population.(144) A number of the most significant areas of association have now been replicated in an independent panel of 1,182 Crohn's disease individuals of European descent using 37 SNPs in tight linkage disequilibrium with SNPs form the index study.(149) ATG16L1 was initially identified in a study of 20,000 nonsynonymous SNPs in 735 patients with Crohn's disease and controls.(78) The SNP (rs2241880) which results in an amino acid change A197T in exon 8 was observed to be significantly less common in the Crohn's disease patients compared to controls, thus conferring protection against Crohn's disease. When multiple regression analysis was carried out using further SNPs in the ATG16L1 locus, virtually all of the disease risk of the gene was explained by this SNP. The association between the protective allele A197T of ATG16L1 and Crohn's disease was also replicated in the USA wide, non- Jewish, European ileal Crohn's disease cohort mentioned above in a further report of their genome wide association study.(150)

In a genome wide scan of the Flemish Crohn's disease population an area of chromosome 5p13 was observed to contain multiple SNPs that were strongly associated with Crohn's

disease.(147) The 1.25- Mb region of chromosome 5 that was identified is a gene desert, although the authors presented some expression data that correlated with variant alleles of the candidate gene prostaglandin receptor EP4 which flanks the area of association. This association was also replicated in the UK wide genome wide scan.(57)

Further combined analysis of the US, UK and Flemish genome wide studies have identified 32 loci that reach statistical significance ($p<10^{-7}$).(151) Eleven of the loci had been previously identified and 21 of the loci were novel.

In the first genome wide scan to be reported in patients with ulcerative colitis, a UK wide study of nonsynonymous SNPs identified a novel susceptibility locus ECM1 that was associated with ulcerative colitis.(152) Interestingly, some of the previous associations with Crohn's disease were replicated- IL-23R, the HLA locus, NKX2.3 and MST1. The autophagy genes ATG16L1 and IRGM were not associated with ulcerative colitis eloquently illustrating the genetic relationship between the two inflammatory bowel diseases. Further genome wide association studies carried out in patients of European ancestry have identified ulcerative colitis risk loci on chromosomes 1p36 and 12q15 that were replicated in independent populations,(153) and a SNP flanking IL-10, a strong candidate gene in ulcerative colitis pathogenesis.(154)

1.10 MICROARRAY GENE EXPRESSION STUDIES

Genome wide microarray expression studies allow a comprehensive picture of gene expression at the tissue and cellular level, thus helping understand the underlying physiological processes. Over the last 10 years microarray publications have had a major impact on gene expression research and have led to a significant number of novel findings.

1.10.1 History of Microarray Technology

Microarray technology was developed from spotted nylon array technology used to identify genomic inserts in bacterial colonies by hybridization with preidentified cDNAs.(155) In the seminal microarray experiment using a two- colour fluorescent pattern of differential gene expression, expression patterns were compared between the root and the leaf of *Arabidopsis.* (156) Twenty seven of the 45 genes examined were found to have significantly different expression levels when the root and leaf were compared and in 26 of these genes the difference in expression was more than 5 fold. Interestingly mRNA from the light regulated chlorophyll a/b binding gene was found to be around 500 times more abundant in the leaf compared to the root. The authors concluded that cDNA microarrays could provide a useful link between human gene sequences and clinical medicine, thus helping in the investigation and understanding of physiological and pathological conditions.

Since 1995 there has been a rapid increase in the number of papers published using microarray technology from 7 in 1995-96 to 139 in 1999 and to 3000 in 2003, (157) with the majority of

these papers using microarrays to measure expression rather than sequencing, re-sequencing or genotyping.(155) The optimism set out by Mark Schena one of the authors of the seminal *Arabidopsis* microarray experiment suggesting that "all human illness can be studied by microarray technology, with the ultimate goal of this work being to develop effective treatments and cures for every human disease by 2050" has yet to be been fully realised, however, a number of exciting and novel observations have been generated.

1.10.2 Basic Microarray Function

Microarray is defined as a tool that allows the simultaneous study of expression of large numbers of genes and initially arrays that contained feature spots of 200μm or smaller were defined as microarrays.(158) To undertake a microarray experiment first RNA is extracted from a biological sample and converted into cDNA or cRNA.(159) The sample is then fluorescently labelled and hybridized on the microarray chip containing known DNA sequences from genes of interest (Figure 1.1). At this point it is worth noting that microarray nomenclature differs from conventional molecular biology nomenclature and the term 'probe' refers to the known sequence which is tethered to the chip, where as the 'target' is the unknown sequence that is in solution and binds to the known 'probe'. Previous conventional techniques have the nomenclature the other way around-'probe' is in solution and 'target' is the known and fixed sequence. The nomenclature was initially reversed by Phimister (160) and this has been carried on by *Affymetrix, (Santa Clara, California)* one of the established microarray manufacturers.(161)

Figure 1.1: Schematic of gene expression analysis using oligonucleotide microarray technology.

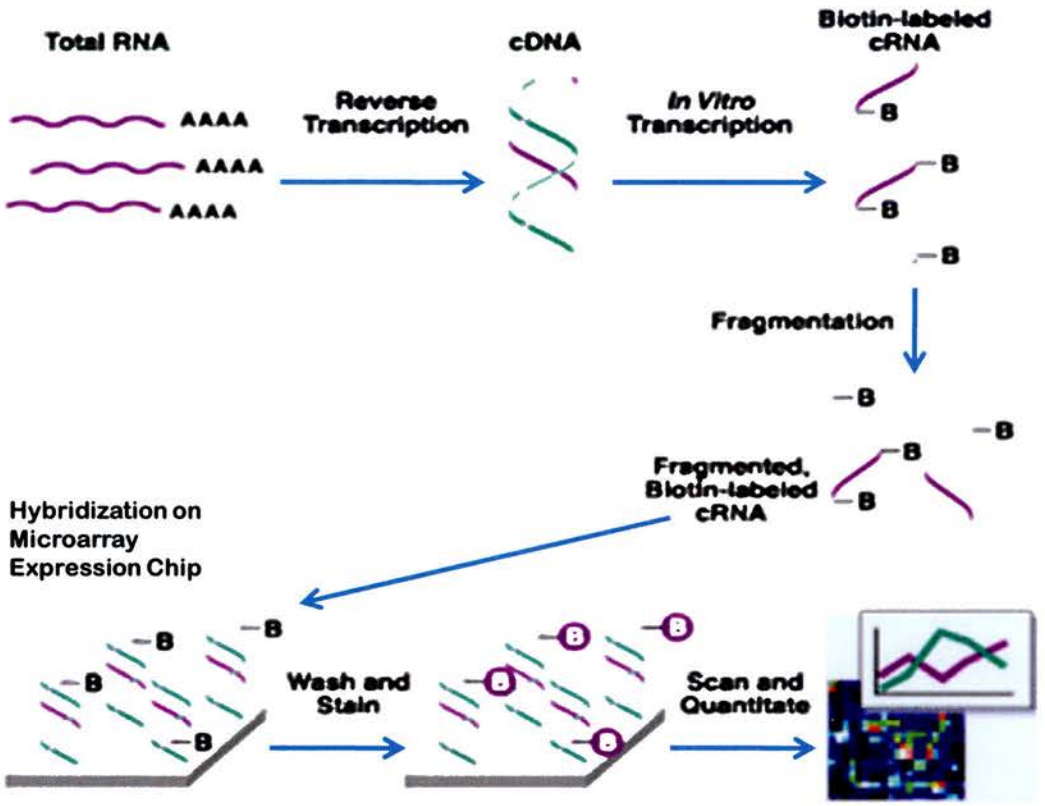


Figure 1.1: Oligonucleotide arrays use short sequence 'probes' that span conserved exons across transcripts of targeted full length genes. Thus, the fluorescent signal from each of the probe pairs from a specific gene are measured and averaged, substantially reducing the risk of cross hybridization and background noise. Figure copied and modified from.(162)

Following the hybridization the microarray chips are washed so that all that remains are the fluorescent 'targets' tethered to the microarray 'probes'. Laser technology is then used to measure the location and intensity of the fluorescent signal from each spot on the microarray

chip and the degree of intensity of fluorescence measured is calibrated to calculate the levels of mRNA expression for each sequence.

1.10.3 Development of Microarray Technology

Genomic research and in particular microarray studies have turned molecular biology from a data poor into a data rich science and to process and manage these large volumes of data new technologies and techniques have been developed.(163) With each step in the miniaturization of microarray technology increasing numbers of sequence probes can be placed on each array chip. In the first study in patients with inflammatory bowel disease in 1997 Heller and colleagues examined the expression of 96 genes whereas the *Agilent* arrays used to generate the data presented in this thesis have 41058 sequences on each chip.(164;165) It is expected that the number of probes on each chip will continue to grow in an exponential manner until the physical limitations imposed by the diffraction of light limit further expansion at around 1μm per feature size.(155)

Attention has also focused on reducing the amount of RNA required for each array chip as RNA from diseased tissue is often only available in small quantities and is difficult to obtain. In microarray studies in patients with inflammatory bowel disease this trend towards the ability to analyse smaller amounts of RNA has been illustrated with initial studies having to use full thickness blocks from surgical resections for analysis,(166;167) where as more recent studies have been able to use endoscopic pinch biopsies.(168) Some initial studies also pooled RNA from a number of diseased samples enabling there to be sufficient quantities of RNA to

be analyzed, however, in many cases this approach resulted in significant volumes of background noise.(157;169)

One of the technologies put forward for enhancing the detection of low levels of RNA from small samples has been the use of longer oligonucleotide probes.(170) *Affymetrix* one of the established microarray manufacturers constructs oligonucleotide arrays containing 25 nucleotides where as *Agilent* probes contain 60 nucleotides.(161;171) Data from Hughes and colleagues examining expression profiles using microarrays fabricated by ink- jet oligonucleotide synthesizer would strongly suggest that 60 nucleotide probes have a better sensitivity and specificity compared to probes with smaller numbers of nucleotides to detect transcripts in complex biological samples.(172) Having larger probes allows the *Agilent* platform to function with only 50ng of RNA for amplified cRNA labelling microarray and this made it the most appropriate platform to use in the present study.(171)

Unpublished work carried out by *Genisphere (Hatfield, Pennsylvania)* has shown that using fluorescently tagged DNA based dendrimers during sample amplification to 'capture' complimentary sequences, microarrays can be carried out on as little as 0.1ng of total RNA which equates to around 10 cells.(170) This technology should better enable researchers to examine cells harvested by laser capture micro- dissection as these analysis have previously been hampered by limited quantity and quality of RNA.(173)

Another area where microarray technology has advanced is in the interrogation of formalin fixed or paraffin embedded tissues, many of which are a number of years old and the RNA has

degraded. By using conventional 60 nucleotide arrays on RNA extracted from these tissues one would expect a 30%- 40% probe success rate compared to a success rate of more than 80% with fresh frozen tissue *(Z Modrusan, Genentech, personal communication)*. However, by using large numbers of tag sequences to quantify amplified primer extension products, samples 20 years old have been analysed.(170)

1.10.4 Experiment design and interpretation

Experimental design and planning are crucial to interpreting the large volumes of microarray data that are generated. This process starts with the collection and phenotyping of the biological material.(174) In the case of inflammatory bowel disease it is paramount to obtain as much phenotypic information about the patients in order to investigate the effect of variables such as disease location and smoking status on gene expression. The scoring of biopsies for the degree of inflammation also helps differentiate the acute inflammatory signal from quiescent disease where subtle changes may yield susceptibility genes.

Microarray technology has been and still is expensive, so a number of studies have been hampered by insufficient replication of results and in some cases the pooling of biological samples, thus generating false positive data.(155) The failure to undertake replication studies has slowly been addressed as the importance of validating results has been realised, technology become cheaper and more funding is made available.

Increasing insight has also been gained into individual probe error distribution and the corrections that are required to model these potential multiplicative errors.(175;176) Different manufacturers have adopted a variety of approaches to ensuring experimental quality and one way is to 'spike- in' known amounts of a positive control at a level roughly corresponding to one copy per cell in order for these sequences to easily detectable and biologically valid.(155) By correcting for these known values comparisons can be made across different data sets. Other platforms have adopted a multiple scan approach, further extending the sensitivity of detection.(171)

1.10.5 Validation of Microarray Results

Comparing different gene expression studies using different microarray formats to validate results has been fraught with difficulty, however, a number of collaborative efforts have now compared different formats and set minimum standards for data presentation.(174;177) When the technology was in its infancy microarray chip manufacturers were reluctant to divulge the exact sequences of their probes so investigators were unable to identify which part of the gene or indeed which splice variant of the gene the probe was targeting.(157) Since 2001 increasing numbers of microchip manufacturers have made their probe sequences readily available on databases such as *ensemble (http://www.ensembl.org/index.html).*

Recent data presented by the Microarray quality control project (MAQC) and the External RNA Controls Consortium (ERCC) have compared gene expression data derived from common samples among different microarray platforms.(178;179) In this study 4 pools of

samples were created- sample A contained 100% Universal human reference RNA (UHRR), sample B contained 100% Human brain reference RNA (HBRR), sample C was composed of 75% UHRR and 25% HBRR and sample D was composed of 25% UHRR and 75% HBRR.(179) Six platforms were tested- *Applied biosystems, Affymetrix, Agilent technologies, GE healthcare, Illumina* and *Eppendorf.*

All the sequence information for the probes was provided by the manufacturers and these were found to represent between 15,429 and 16,990 genes. To simplify comparison one probe was chosen to represent each gene and genes were selected if they were represented on each platform resulting in 12,091 probes matching 12,091 common reference sequences from 12,091 genes. Overall the median coefficient of variation ranged from 10% to just over 20% which was not dramatically higher than the replicate values calculated on the same platform. The results displayed a high level of intra-platform consistency across different test sites as well as a high level of intra- platform concordance with comparing differentially expressed genes. Three of the platforms- the *Affymetrix, Illumina* and the *Agilent* platforms also displayed a correlation of greater than 0.9 with TaqMan assays of 450 of the examined genes adding further validity to the results.

In a parallel publication, one, two and three colour microarray platforms were investigated and found to have high correlation coefficients suggesting that the colour number need not be a factor in determining microarray design.(180) In a further study by Canales and colleagues a high correlation was again observed between microarray data from three platforms and 997 TaqMan assays, 205 real time PCR assays and 244 QuantiGene assays.(181) Taken together

these data validate the use of carefully designed and executed trials using microarray technology to quantatively characterise gene expression.

1.10.6 Statistical analysis of microarray data

Microarray technology has driven forward an experimental method that is based on a hypothesis free, unbiased sample screening approach and to manage these novel data sets new data interpretation techniques have been required.(182) A number of initial microarray studies only published fold changes without p values and confidence intervals, however, as the field has matured more rigorous statistical techniques are being used.(183)

As thousands of genes are looked at simultaneously correction for multiple comparisons is important as if 20000 genes are investigated for expression changes with a p threshold of 0.01, 200 genes would by differentially expressed by chance alone. A *Bonferroni* correction corrects the p value for the number of comparisons in each experiment, however, many would argue that the *Bonferroni* correction is too conservative a correction and true biological differences will be missed.(163) Less conservative corrections for multiple comparisons have been proposed and these probably strike a more realistic balance between eliminating false positives and observing true biological differences.(184)

To correct for multiple hypothesis testing, Storey and colleagues have proposed a method for calculating a q-value for each tested feature on the microarray chip.(185) The q-value estimates significance in terms of the false discovery rate rather than the false positive rate.

From the q-values a false discovery rate for each individual analysis can be calculated and a false discovery rate of less than 5% is considered satisfactory.

1.10.7 Gene Ontology

Gene ontology is the categorization of all of the genes in a field of knowledge into functional pathways showing the relationships between each gene. By analyzing genes that are differentially regulated in microarray experiments and plotting their biological pathways and function, networks of interacting genes/ proteins can be identified and potentially targeted. To be able to plot and define these pathways, details about individual protein function and interaction need to be understood and annotated and this has involved a considerable amount manual labour.(186) Automated systems involving entering genes discussed in publications from *PubMed* to ontology data sets are being developed and an example of a gene ontology map is shown in Figure 1.2.

With increasingly large numbers of sequences that represent genes being recognised by ontology software packages, a more meaningful interpretation of the data can be gathered and this type of analysis provides a conceptual framework for semantic representation of textual information.(187)

Figure 1.2: A Gene ontology map involving a subset of differentially regulated genes in endoscopic biopsies of patients with inflammatory bowel disease.

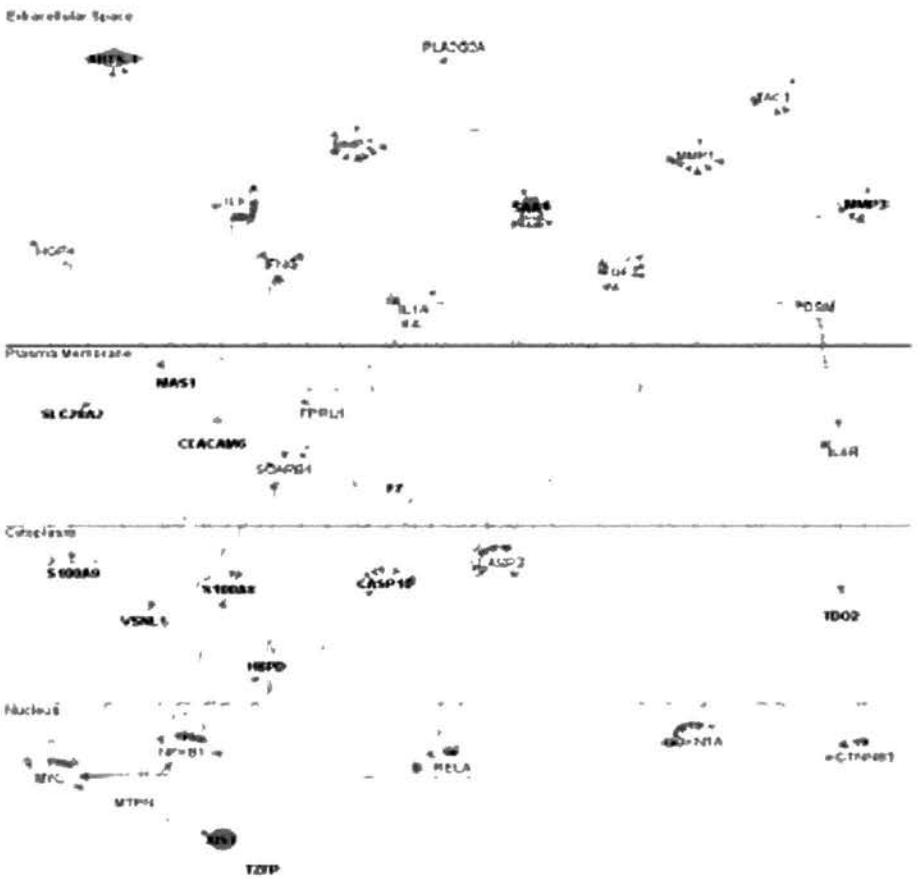


Figure 1.2: Upregulated genes are shown in a spectrum of pink where there is a small level of upregulation to red where there is a high degree of upregulation. Downregulated genes are shown in a spectrum of light green where there is a small level of downregulation to dark green where there is a high degree of downregulation. No change in expression was noted in genes shown in white. The area of the cell in which the protein products function is also illustrated. Solid lines between genes illustrate confirmed interactions in the direction of the arrow and hatched lines between the genes show probable interactions in the direction of the arrow. This map was created using the *Ingenuity Pathway analysis software (Redwood City, California).*

1.10.8 Future Implications

As the cost of microarray technology declines and results generated become more robust microarrays will increasingly become available as a routine diagnostic test in clinical medicine. Also with the miniaturization of the technology, smaller samples including those captured by laser micro- dissection will be analysed with greater sensitivity. Microarrays are increasingly being used for genotyping with simple discrimination arrays used to identify SNPs in the DNA.(188) Genome wide epigenetic analysis and splice variant analysis are two further fields where microarray based techniques are being developed.

1.11 CLINICAL DATA GENERATED BY MICROARRAY STUDIES

Probably the most provocative and clinically relevant data generated by microarray has been in the field of cancer research. An example of the potential of microarray was demonstrated by Alizadeh and colleagues who carried out microarray analysis on 96 samples ranging from normal lymph node tissue to a variety of different non-Hodgkin's lymphoma.(189) By comparing expression analysis in a 'Lymphochip' panel of 3,186 genes between patients with diffuse large B- cell lymphoma (DLBCL) prior to treatment the investigators were able to group these patients into two discrete groups- germinal centre DLBCL and activated DLBCL. A subset of these markers were followed up and validated by PCR. When the clinical progress of these two groups of patients were examined the germinal cell DLBCL had a higher five year survival than the activated DLBCL (75% n=25 versus 16% n=37 respectively, $p < 0.01$) prompting the authors to suggest that the DLBCL category harboured two different diseases and that expression profiling in combination with clinical indicators would help manage therapeutic intervention in these patients.

Gene expression profiles have also been successfully used to predict prognosis in 295 patients with breast cancer,(190) however, a meta- analysis of 84 studies found that DNA microarrays had a variable prognostic performance in measuring prognosis in a number of different cancers.(191) In this meta-analysis, larger studies with appropriate clinical design and larger probe sets yielded better prognostic information than smaller studies.

Because of the relative ease of biopsy of the gastrointestinal tract there have been a number of microarray studies in gastrointestinal cancer and coeliac disease. Differential expression of duodenal biopsies in patients with coeliac disease who were untreated, patients on a gluten-free diet and healthy controls have led to the discovery of a number of candidate genes that are under further investigation.(192;193)

1.11.1 Microarray Studies in Patients with Inflammatory Bowel Disease

In 1997 microarrays were used to compare synovial tissue obtained from patients with advanced rheumatoid arthritis undergoing arthroplasty or remedial synovectomy with inflamed bowel of patients undergoing surgery for Crohn's disease.(194) Ninety six candidate genes were selected and their expression was compared in the rheumatoid arthritis and the Crohn's disease tissue. As expected many inflammatory genes were commonly expressed between the two diseased tissues and the authors then went on to investigate several genes that were expressed preferentially in the rheumatoid arthritis tissue as opposed to the Crohn's disease tissue. In the three Crohn's disease samples that were microarrayed IL-1 converting enzyme, tissue inhibitor of metalloproteinase 1 (TIMP-1) and Migration inhibitory factor (MIF) had notably increased expression.

In 2000 Dieckgraefe and colleagues published a study using microarray technology to compare eight patients who had ulcerative colitis and were having colectomies for disease refractory to medical management and a control group.(166) The severity of the ulcerative colitis was assessed by taking a paired biopsy and having a pathologist assess it for disease

severity. Because these samples reflected the more severe end of the disease activity spectrum one specimen was taken from uninvolved mucosa of one patient with ulcerative colitis having a colectomy. The control group was made up of 3 non inflamed samples, resections for adenocarcinoma, diverticular abscess and diverticulitis. Four patients formed the inflamed control group, one patient with an inflamed colon as a result of a rectal prolapse and three patients with Crohn's disease-1 fistulating, 1 ileal and 1 Crohn's colitis.

These samples were then microarrayed and 6500 genes were analysed. The results confirmed increases in a number of genes previously implicated in the pathogenesis of ulcerative colitis (IL1, IL1 RA and IL8) and suggested that multiple members of the chemokine subfamily may play a role in disease pathogenesis.

In 2001 Lawrance and colleagues assessed gene expression in surgically resected colonic specimens from six patients with Crohn's disease, six patients with ulcerative colitis and six control samples.(167) The inflammatory bowel disease samples had their diagnosis confirmed endoscopically and a pathologist selected biopsies from areas of moderately severe inflammation in a blinded and random fashion. The control samples were taken from 4 colon cancers, one diverticular disease and one caecal polyp and tissue was taken at least 10cm away from the area of pathology. The RNA was extracted from the full thickness colonic tissues. The respective groups of samples were pooled and microarrayed.

7070 genes were examined and a three fold change was judged to be a significant threshold for differential gene expression. 170 genes were differentially expressed in ulcerative colitis

and Crohn's disease with almost an equal number up regulated and down regulated. 20% of the differentially regulated genes were common to both forms of inflammatory bowel disease. The locations of these genes were mapped and several were found to lie within IBD2 locus which is located on chromosome 12. Comparing ulcerative colitis and Crohn's disease, in the ulcerative colitis samples there was over expression of HLA 1 transcripts and in Crohn's disease samples were associated with an over expression of antimicrobial defensins.

Endoscopic investigation of inflammatory bowel disease with the ability to take pinch mucosal biopsies has allowed investigators to microarray tissue from a larger range of patients, encompassing those with less severe disease compared to patients who have required resectional surgery. Langman and colleagues used microarrays to analyse biopsy specimens from macroscopically non affected areas of the colon and terminal ileum.(168) Biopsies were taken from nine patients with Crohn's disease, 18 patients with ulcerative colitis and 14 control patients who were having colonoscopic surveillance as a result of having a family history of colon cancer. Tissue samples from the inflammatory bowel disease patients were taken from at least 10cm away from pathological areas of inflammation. 22,283 genes were analyzed and the samples were pooled from four patients in each group. Terminal ileal and transverse colon samples were pooled to represent gene expression across the whole bowel.

Interestingly genes which were involved in cellular detoxification and biotransformation (pregnane X receptor and MDR1) were significantly downregulated in the colon of patients with ulcerative colitis, however, there was no change in the expression of these genes in the biopsies from patients with Crohn's disease.

In a study of 24 patients with ulcerative colitis Okahara and colleagues investigated the difference in gene expression between endoscopic biopsies taken from inflamed and non inflamed areas using a 1300 gene cDNA microarray.(195) Using a stringent three fold difference in expression the authors found five genes- (Migration inhibitory factor- related protein 14 (MRP14), growth-related oncogene gamma (GROγ) and serum amyloid A1 (SAA1) were upregulated where as TIMP1 and elfin were down regulated in the inflamed biopsies when compared to the non-inflamed biopsies. These results were confirmed using real time PCR.

Costello and colleagues examined the expression of 33792 sequences in endoscopic sigmoid colon biopsies obtained from 11 healthy controls, 10 patients with Crohn's disease and 10 patients with ulcerative colitis.(196) A number of sequences representing novel proteins were differentially regulated and in silico analysis suggested that these proteins had putative functions related to disease pathogenesis-transcription factors, signaling molecules and cell adhesion.

Chemokines are known to determine inflammatory leukocyte recruitment and retention and Puleston and colleagues investigated expression of 41 chemokines and 21 chemokine receptors in 101 endoscopic biopsies from patients with Crohn's disease, ulcerative colitis and histologically normal controls.(197) Chemokines CXCLs 1-3 and 8 and CCL20 were upregulated in active colonic Crohn's disease and ulcerative colitis when compared to uninflamed inflammatory bowel disease biopsies and controls. Real time PCR showed

increased expression of the receptors of the chemokines that were upregulated in the inflamed biopsies (CXCR1, CXCR2 and CCR6).

Microarray analysis has also been used to analyse peripheral blood mononuclear cells in patients with inflammatory bowel disease. To attempt to identify different expression signatures from patients with Crohn's disease and patients with ulcerative colitis Burczynski and colleagues studied peripheral blood mononuclear cells from 59 patients with Crohn's disease and 26 patients with ulcerative colitis.(198) Twelve genes were identified that distinguished Crohn's disease patients from ulcerative colitis patients with a high degree of accuracy and the authors went on to speculate that looking at expression of these genes would be a useful way to determine the diagnosis in patients with indeterminate colitis.

In a further microarray experiment looking at peripheral blood mononuclear cells in Crohn's disease patients who were homozygote or compound heterozygote for the three common NOD2/CARD15 variants- Arg702Trp, Gly908Arg and Leu1007fsinsC, and comparing these samples to Crohn's disease patients who were wild type for these NOD2/CARD15 variants, a globally blunted transcriptional response was observed in the case of the NOD2/CARD15 variants following the addition of muramyl dipeptide.(199)

These data generated from surgical resections, endoscopic biopsies and peripheral blood mononuclear cells from patients with inflammatory bowel disease illustrate how useful microarrays have been in the investigation and understanding of inflammatory bowel disease. They also show how as microarray technology has developed with increasingly large numbers

of sequences being analysed with greater sensitivity and specificity in increasingly small biological samples.

1.12 OTHER MOLECULAR IDENTIFICATION STRATIGIES

1.12.1 Proteomics

Proteomic analysis is another new investigative tool that has not been used to any large extent in the field of inflammatory bowel disease. Proteomics refers to the study of the total protein content of cells.(200) This is achieved by combining the techniques of protein electrophoresis and mass spectrometry. Expression profiles of proteomes may be generated from samples of serum or secreted fluid, and these may be able to differentiate disease progression, response to therapy and identify novel therapeutic targets.

Preliminary data has been generated in our unit comparing serum of patients with severe ulcerative colitis who responded to corticosteroid therapy and matched patients who were resistant to and failed corticosteroid therapy.(201) Proteomic profiles of corticosteroid resistant and responsive groups were significantly different with 12 proteins up-regulated and 7 proteins down-regulated in the corticosteroid resistant group. These results suggest that protein profiling may be useful in predicting patient response to corticosteroid therapy and identification of these proteins is currently underway.

1.12.2 Yeast Two-Hybrid Screening

The yeast two-hybrid assay is an elegant means of investigating protein-protein interactions, which have become increasingly important in our understanding of biological systems and

pathways. The yeast two-hybrid system can also be used to characterize interactions already known to occur, thus helping delineate the protein domains responsible for the interaction and the environmental conditions involved.(202)

The yeast two-hybrid assay is performed in the budding yeast, *S cerevisiae* using two fusion proteins: the target protein of interest, known as 'bait' is fused to a DNA binding domain attached to its N-terminus. The second protein, the 'prey' is fused to an activation domain. If the bait protein interacts with the prey, these bring the binding domain and the activation domain of transcriptional activator together, which in turn switches on the expression of the reporter genes. The reporter genes are constructed to allow growth of the yeast in a selective medium when the interaction occurs. When investigating 'protein- protein' interactions, a single bait protein is used to search for interaction with a library of proteins fused to the activation domain. The choice of library is determined by the tissue of interest, e.g. intestinal cell library for inflammatory bowel disease.(203)

Schizophrenia is a disease of polygenic genetic susceptibility where the yeast two hybrid system has been successfully used to identify candidate genes. The disrupted-in-schizophrenia 1 (DISC1) gene which was identified in 2000 (204) and confirmed in other cohorts,(205;206) was used as bait. DISC1 encodes a novel protein of unknown function and full-length human DISC1 protein was used to screen human adult and foetal brain libraries for interacting proteins, using the yeast two-hybrid system.(207) Twenty-one proteins from a variety of locations were identified implicating DISC1 in several aspects of central nervous

system signalling and confirming data from other yeast two-hybrid scans using DISC1 as bait.(208;209) From these data the authors were able to identify a number of potential DISC1 interactions and they were able to speculate that DISC1 one may be at the centre of an extensive protein interaction network.

In inflammatory bowel disease, Barnich and colleagues used NOD2/CARD15 gene as 'bait' to screen a bone marrow library and they identified GRIM19, a protein with homology to the NADPH dehydrogenase which interacted with endogenous NOD2.(210) GRIM19 is required for NFκB activation following NOD2 mediated recognition of bacterial muramyl dipeptide and the authors hypothesised that GRIM19 is a key component of the CARD15/NOD2 signalling pathway which currently remains under detailed investigation.(49-51)

CHAPTER 2

AIMS AND OBJECTIVES

From the discovery of the NOD2/CARD15 gene to the era of genome wide association studies, genetic studies in inflammatory bowel disease have advanced at a rapid rate. The plan for this thesis was created in 2004 to follow up the *Nature Genetics* publications of Stoll and Peltekova and colleagues,(95;107) along side investigating genome wide expression in endoscopic colonic and terminal ileal biopsies of patients with inflammatory bowel disease using microarray technology. As such the genetic data presented in this thesis must be seen in the context of this fast moving field. The primary aims were:

1) To investigate gene expression profiles in human colonic and terminal ileal biopsies using microarray technology in a well phenotyped cohort of patients with Crohn's disease, ulcerative colitis and a control group. The role in disease pathogenesis of differentially expressed genes was investigated along with expression of candidate genes identified by genome wide association study and cell lineage analysis.

2) To investigate the contribution of polymorphisms in the IBD5 locus in determining susceptibility and disease phenotype in Crohn's disease and ulcerative colitis in the Scottish population.

3) To investigate the contribution of variants rs1248696 (113A) and rs2289311 in the candidate gene DLG5 in determining genetic susceptibility to Crohn's disease and ulcerative colitis in the Scottish population.

4) To examine the colonic and terminal ileal expression of genes in the IBD5 locus and Drosophila Discs Large Homolog 5 (DLG5) in health and in inflammatory bowel disease.

CHAPTER 3

MATERIALS AND METHODS

EQUIPMENT AND REAGENTS

3.1.1 Equipment

ABI Techne Touchgene PCR Gradient machine

Agilent feature extraction software (version 7.5) (Agilent Technologies).

Agilent G2505B model slide scanner (Agilent Technologies).

Agilent whole human genome oligo microarray chips G4112A.

Endoscopic biopsy forceps (Olympus Endoscopy, Tokyo, Japan)

Endoscope and endoscopy stack (Olympus Endoscopy).

Endoscopic scope guide (Olympus Endoscopy).

Graph Pad Prism software (San Diego, CA)

Hybridization rotator oven for Agilent Microaray Chambers (Agilent Technologies).

Ingenuity Software (Mountain View, CA)

Microsoft Office XP/ 2007 software (Microsoft, Seattle).

Minitab software (Coventry, UK).

NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, Delaware)

Rosetta Resolver software (Rosetta Inpharmatics, Seattle).

Stratagene model MX4000 (La Jolla, Ca, USA).

TaqMan® ABI Fast Real Time PCR 7600 HT (Applied Biosystems)

TissueRuptor (Quiagen, Valencia, CA)

3.1.2 Reagents

Micro total RNA isolation from animal tissues kit 74004 (Quiagen)

RNeasy MinElute spin columns
Buffer RLT
Buffer RW1
Buffer RPE
RNase free water
Buffer RDD
RNase free DNase
Carrier RNA, poly A

Low RNA input fluorescent linear amplification kit 5184-3523 (Agilent Technologies)

T7 promoter primer
5X first strand buffer
0.1M DDT
0.1 mM dNTP mix
MMLV RT
RNaseOUT
4X Transcription Buffer
NTP mix
Inorganic Pyrophosphatase
T7 RNA polymerase
50% polyethylene glycol
CTP
Random Hexamers
dNTP
RNAse I 'A'

RNeasy Mini Kit 74104 (Qiagen)

RNeasy mini spin columns
Buffer RLT
Buffer RW1
Buffer RPE
RNase free water

Agilent stabilization and drying solution protocol 5185- 5976 (Agilent technologies).

Gene Expression Wash buffer 1
Gene Expression Wash buffer 2
DNase/ RNase free water

100% Ethanol

MessageAmp™ II aRNA Amplification Kit AM1751 (Ambion technologies, Austin, TX).

T7 Oligo (dT) primer
ArrayScript
RNase inhibitor
10X First strand buffer
dNTP Mix
10X Second strand buffer
DNA polymerase
RNase H
T7 Enzyme Mix
T7 10X Reaction buffer
T7 ATP
T7 CTP
T7 GTP
T7 UTP
Control RNA

3.2 MICROARRAY METHODS

3.2.1 Overview

The microarray study was set up so that the patients were recruited at the Western General Hospital in Edinburgh and then the snap frozen biopsies were sent for microarray analysis to Genenetch Inc, in South San Francisco. For one year, all the patients who were attending for colonoscopy at the Western General Hospital in Edinburgh were written to if they had existing inflammatory bowel disease or symptoms consistent with a new diagnosis of inflammatory bowel disease. During this year 67 patients with ulcerative colitis, 53 patients with Crohn's disease and 31 control patients who were undergoing colonoscopy for other indications were recruited. This gave us a data set that was at least three times larger than any previously published microarray series in inflammatory bowel disease.

3.2.2 Recruitment of Patients for Endoscopic Biopsy Collection

The demographics of the ulcerative colitis and Crohn's disease patients are described in detail in the subsequent chapters. All inflammatory bowel disease patients attended the clinic at the Western General Hospital, Edinburgh and the diagnosis of inflammatory bowel disease adhered to the criteria of Lennard-Jones.(211)

Phenotypic data were collected by interview and case-note review and comprised of demographics, date of diagnosis, Vienna classification of disease location, disease behaviour

and progression,(115) extra-intestinal manifestations, surgical operations, current medication, smoking history, family history and ethnicity. Data were recorded on a proforma. At the time of colonoscopy patients symptoms were evaluated using the simple clinical colitis activity index (SCCAI) (212) for patients with ulcerative colitis and the Harvey Bradshaw Index for patients with Crohn's dsease.(213) Patients with inflammatory bowel disease were defined as 'new diagnosis' if the colonoscopy took place at the time of their index presentation and they had less than 24 hours of oral/IV therapy.

Eleven of the controls were male, 20 were female and they had a median age of 43 at the time of endoscopy. Six of the controls had normal colonoscopies for colon cancer screening, 9 controls had symptoms consistent with irritable bowel syndrome and had a normal colonoscopic investigation and 7 patients had a colonoscopy for an other indication and histologically normal biopsies were obtained. Eight control patients had abnormal inflamed colonic biopsies (1 pseudomembranous colitis, 1 diverticulitis, 1 amoebiasis, 2 microscopic colitis, 1 eosoinophilic infiltrate, 2 scattered lymphoid aggregates and a history of gastroenteritis).

3.2.3 Microarray Ethical Approval

Written informed consent was obtained from all patients. Lothian Local Research Ethics Committee approved the study protocol: REC 04/S1103/22.

3.2.4 Biopsy Collection

Anatomical location was confirmed by operator experience, distance of endoscope insertion and endoscope configuration using a Scope Guide ™. Paired biopsies were taken from each anatomical location. One biopsy was sent for histological examination and the other was snap frozen in liquid nitrogen for RNA extraction. Each biopsy was graded histologically into those with no evidence on inflammation, biopsies with evidence of chronic inflammation and a chronic inflammatory cell infiltrate and those with acute inflammation and an acute inflammatory cell infiltrate. Two hundred and forty five paired inflammatory bowel disease biopsies and 76 paired control biopsies were collected. The number of paired biopsies from each anatomical location are shown in Table 3.1.

Table 3.1: The location and number of biopsies in ulcerative colitis, Crohn's disease patients and controls

	Ulcerative Colitis (n =67)	**Crohn's disease (n=53)**	**Controls (n =31)**
Total number of paired biopsies	139	106	76
Terminal Ileum	4	16	6
Ascending colon	33	20	17
Descending colon	35	30	23
Sigmoid colon biopsies.	57	33	27
Removed from analysis	10	7	3

3.2.5 RNA Isolation

RNA was isolated from the biopsies by Jennine Cornelius and Colin Noble at Genentech. The biopsies weighed between 0.2mg and 16.5mg with a mean weight of 5.5mg. The biopsies

were disrupted using a TissueRuptor in 350µl of Buffer RLT for 30 seconds. The supernatant was then centrifuged for 3 minutes and 70% ethanol was added. The sample was placed in an RNeasy MinElute spin column and centrifuged for 15 seconds and the flow through was discarded. 350µl of RW1 buffer was added. 10µl of DNase was then added to 70µl of buffer RDD and this was added to the samples and left for for 15 minutes. The samples were again centrifuged for 15 seconds and the follow through was discarded. The samples were washed with 350µl of buffer RW1 and the follow through was discarded. 500µl of buffer RPE was then added and the samples and they were centrifuged for 15 seconds and the follow through discarded. 80% ethanol was added and the samples were centrifuged for 2 minutes and the follow through was discarded. The MinElute column was then removed and centrifuged for a further 5 minutes. 14µl of RNase free water was added and centrifugation was carried out for 1 minute to elute the RNA. To evaluate purity and integrity 1µl of total RNA was assessed from each sample with the Agilent technologies 2100 bioanalyzer using the Pico LabChip reagent set.

3.2.6 Low Input Fluorescent Linear Amplification

The amplifaication was carried out by Karen Toy and Colin Noble at Genentech. A schematic of the Agilent low input fluorescent linear amplification protocol is shown in Figure 3.1. One 1µg of RNA was added to a centrifuge tube along with 5µl of T7 promoter primer and the total volume was brought up to 11.5µl with nuclease free water. The primer was denatured by incubating at 65°C for 10 minutes. The samples were then placed on ice. 8.5µl of the cDNA

master mix were then added to each sample and the samples were incubated in a circulating water bath at 40°C for 2 hours (Table 3.2).

Table 3.2: cDNA Master Mix

Reagent Stock	Amount (µl)
5X-1st Strand cDNA buffer	4.0
0.1 M DTT	2.0
10 mM dNTP mix	1.0
MMLV RT	1.0
RNaseOUT	0.5
total volume (µl)	8.5

The samples were then heated to 65°C for 15 minutes to inactivate the MMLV RT and the samples were then placed on ice for 5 minutes and then briefly centrifuged.

To each sample tube 2.4µl of cyanine 3-CTP (10mM) or cyanine 5-CTP (10mM) was added minimizing light exposure and the samples were warmed to 40°C for 1 minute and then vortexed briefly. At room temperature 57.6µl of transcription master mix was added and the samples were left in a covered water bath for at 40°C for 2 hours (Table 3.3).

Figure 3.1: ***Agilent*** **low input RNA amplification process**

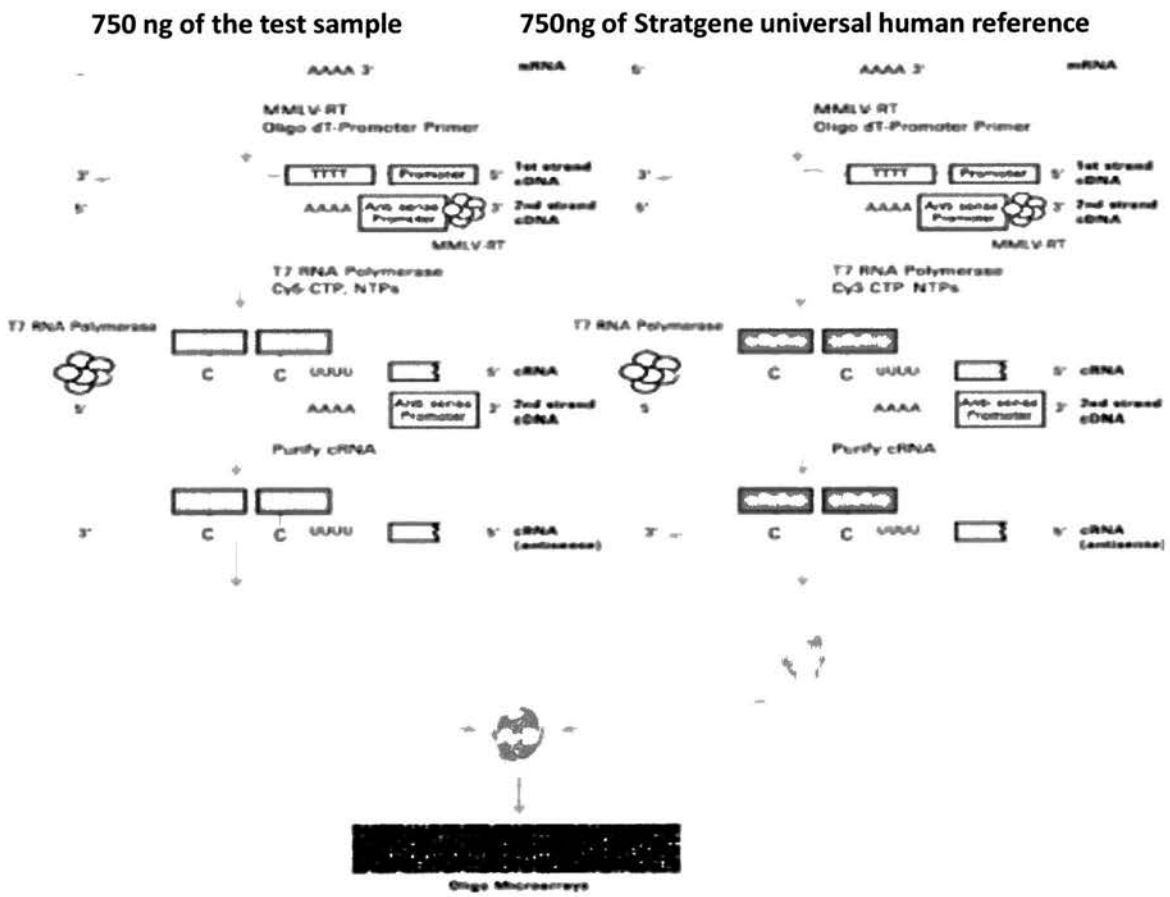


Figure 3.1: A schematic representation of the low input RNA amplification process for cRNA oligo microarray analysis. Copied and modified from www.chem.agilent.com/scripts/ generic.asp?lpage =9234&indcol= N&prodcol=Y

Table 3.3: Transcription Master Mix

Reagent Stock	Amount (µl)
H2O	15.3
4x Transcription buffer	20.0
0.1 M DTT	6.0
NTP mix	8.0
50% PEG	6.4
RNaseOUT	0.5
Inorganic Pyrophosphatase	0.6
T7 RNA Polymerase	0.8
Cy5 or Cy3	2.4
total volume (µl)	60.0

3.2.7 Purification of Amplified cRNA

The purification was carried out by Karen Toy at Genentech. 20µl of nuclease free water was added to each sample to make a total volume of 100µl. To each sample 350µl of buffer RLT was then added and the samples were mixed thoroughly. 250µl of 100% ethanol was added, the sample was mixed thoroughly and 700µl was placed in a RNeasy mini column. The samples were centrifuged for 30 seconds at 9447 times gravity (x g) and the follow through was discarded. 500µl of buffer RPE was then added and the samples were centrifuged for 30 seconds at 9477 x g and the follow through was discarded. 500µl of buffer RPE was again added and the samples were centrifuged for 60 seconds at 9477 x g and the follow through was discarded. The RNeasy column was then transferred to a new 1.5ml collection tube and 30µl of nuclease free water was added and the samples were left for 60 seconds.

Centrifugation was then carried out at 9477 x g for 30 seconds and the follow through was kept. 30µl of nuclease free water was again added and the samples were left for 60 seconds.

Centrifugation was then carried out at 9477 x g for 30 seconds and the follow through was kept making a total final volume of around 60µl which was stored at -80°C. 1µl and of amplified cRNA was quantified using the NanoDrop ND-1000 Spectrophotometer.

3.2.8 Agilent Oligonucleotide Microarray Hybridization

The hybridization was carried out by Karen Toy and Colin Noble at Genentech. 750ng of Stratgene universal human reference labelled with Cy-3 and 750 ng of the test sample labelled with Cy-5 were fragmented for 30 minutes at 60°C in the dark using the target preparation mix in Table 3.4. 250µl of hybridization buffer was then added to terminate the fragmentation reaction.

Table 3.4 Target Hybridization Mix

Components	**1x44K**
cyanine 3-labeled cRNA	750 ng
cyanine 5-labeled cRNA	750 ng
10x Control Targets	50 µl
Nuclease-free H2O	up to 240 µl
25X Fragmentation buffer	10 µl
total volume	250 µl

The samples were then centrifuged at 9477 x g and placed on ice before 490µl was loaded onto the Agilent whole human genome oligo microarray slide G4112A. The samples were hybridized for 18 hours at 60°C with constant rotation.

3.2.9 Stabilization, Drying and Scanning Protocol

The stabilization, drying and scanning protocol was carried out by Karen Toy and Colin Noble at Genentech. Following hybridization the slides were quickly placed in wash 1 (Agilent stabilization and drying protocol) and covered for 10 minutes at room temperature. They were then moved into wash 2 on ice for 5 minutes and finally into wash 3 at room temperature for 5 seconds. The slides were loaded into the agilent scanner.

Microarray slides were scanned using the Agilent G2505B model. Expression signals were calculated using the Agilent feature extraction software (version 7.5). The scan settings are shown in table 3.5.

Table 3.5 Agilent Scan Settings

	Array
Scan Settings	**1x44K**
Scan region (Scan Area)	61x21.6 mm
Scan resolution (μm)	10
5μm scanning mode eXtended Dynamic range	
Dye Channel	Red&Green
Green PMT	100%
Red PMT	100%

3.3.1 Real Time PCR

Real time PCR was undertaken by Jennine Cornelius at Genentech. It was carried out on 8 genes in a cohort of the ulcerative colitis and control biopsies- SAA1, IL8, DEFA5, DEFA6, MMP3, MMP7, S100A8 and TLR4 and 5 genes in the terminal ileal Crohn's disease biopsies-

SAA1, IL8, DEFA5, DEFA6 and MMP3. Prior to RTPCR analysis 1 RNA amplification cycle was carried out using the MessageAmp™ II aRNA Amplification Kit. 5µg of total RNA was added to 6µl of RNA free water with 1µg of T7 Oligo (dT) primer. This was incubated for 10 minutes at 70°C and then centrifuged. The reverse transcription master mix was then prepared (Table 3.5) and the samples were incubated for 2 hours at 42°C, briefly centrifuged and then placed on ice.

Table 3.5: Reverse Transcription Master Mix for Single Amplification Cycle

Reagent Stock	Amount µ1
10X First Strand Buffer	2
dNTP Mix	4
RNase Inhibitor	1
Reverse Transcriptase	1

Reverse transcription PCR was then performed on 50ng of RNA using Stratagene model MX4000. TaqMan primers and probes were manufactured in house at Genentech. (Table 3.6) PCR conditions comprised of 48°C for 30 minutes, 95°C hold for 10 minutes, followed by 40 cycles of 30 second 95°C melt and 1 minute 60°C anneal/extend. (Table 3.6) Absolute quantification of product was calculated by normalizing to RPL19. Results were analysed using Minitab software (Coventry, England).

Table 3.5: TaqMan primers and probes

Gene		Sequence
SAA1	forward	agcgatgccagagagaata
	Reverse	ggaagtgattggggtctttg
	Taq	ctttggccatggtgcggagg
IL8	forward	actcccagtcttgtcattgc
	Reverse	caagtttcaaccagcaagaa
	Taq	tgtgttggtagtgctgtgttgaattacgg,
DEFA5	forward	gctacccgtgagtccctct
	Reverse	tcttgcactgctttggtttc
	Taq	tgtgtgaaatcagtggccgcct
DEFA6	forward	agagctttgggctcaacaag
	Reverse	atgacagtgcaggtcccata
	Taq	cacttgccattgcagaaggtcctg
MMP3	forward	aagggaacttgagcgtgaat
	Reverse	gagtgcttccccttctcttg
	Taq	ggcattcaaatgggctgctgc
MMP7	forward	cacttcgatgaggatgaacg
	Reverse	gtcccatacccaaagaatgg
	Taq	ctggacggatggtagcagtctaggga
S100A8	forward	ttgaccgagctggagaaag
	Reverse	tcaggtcatccctgtagacg
	Taq	tccctgataaaggggaatttccatgc
TLR4	forward	agagccgctggtgtatcttt
	Reverse	ccttctgcaggacaatgaag
	Taq	tggcagtttctgagcagtcgtgc

3.3.2 In Situ Hybridization for Defensin Alpha 5

Navneet Pal and Fiona Zhong (Core Labs, Genentech Inc) designed PCR primers to amplify a 318 bp fragment of human defensin alpha 5 spanning from nt 55-372 of NM_021010 (upper- 5' catccttgctgccattct and lower- 5' ggaccttgaactgaatcttgc). Primers included extensions encoding 27-nucleotide T7 or T3 RNA polymerase initiation sites to allow in vitro transcription of sense or antisense probes, respectively, from the amplified products. Surgical biopsies were fixed in 10% neutral buffered formalin and paraffin-embedded. Sections 5 μm

thick were deparaffinized, deproteinated in 10 ug/ml Proteinase K (Amresco) for 45 minutes at 37 °C, and further processed for *in situ* hybridization as previously described.(214) ^{33}P-UTP labelled sense and antisense probes were hybridized to the sections at 55°C overnight. Unhybridized probe was removed by incubation in 20μg/ml RNAse for 30 min at 37°C, followed by a high stringency wash at 55°C in 0.1 X SSC for 2 hours and dehydration through graded ethanols. The slides were dipped in NTB nuclear track emulsion (Eastman Kodak), exposed in sealed plastic slide boxes containing dessicant for 4 weeks at 4°C , developed and counterstained with hematoxylin and eosin.

3.3.3 Immunohistochemistry for Rabbit Anti-Human Lysozyme and Rabbit Anti-Human Defensin Alpha 6

Sreedevi Chalasani (Core Labs, Genentech Inc) formalin fixed paraffin embedded tissue sections and deparaffinized them prior to quenching of endogenous peroxidase activity (KPL, Gathersburg, MD) and blocking of avidin and biotin (Vector, Burlingame, CA). Sections were blocked for 30 minutes with 10% normal goat serum in PBS with 3% BSA. Tissue sections were then incubated with primary antibodies for 60 minutes, biotinylated secondary antibodies for 30 min, and incubated in ABC reagent (Vector) for 30 minutes followed by a 5 minute incubation in Metal Enhanced DAB (Pierce, Rockford, IL). The sections were then counterstained with Mayer's hematoxylin. Primary antibodies used were rabbit anti-human lysozyme at 5.0 μg/ml (Dako, Carpinteria, CA) and rabbit anti-human defensin 6 alpha at 5.0 μg/ml (Alpha Diagnostics, SanAntonio, TX). Secondary antibody used was biotinylated goat anti-rabbit IgG at 7.5 μg/ml (Vector) Defensin 6 alpha staining required pre-treatment with

Target Retrieval High pH (Dako, Carpenteria,CA) at 99C for 20 minutes, lysozyme staining did not require pretreatment. All other steps were performed at room temperature.

3.3.4 Statistical Analysis

The microarray expression values were normalized using the *Stratagene* Universal Human Reference. For quality control purposes the distribution of log intensities for each sample were plotted and outlying samples (greater than 2 standard deviations from the mean) were excluded from the subsequent analyses- 10 ulcerative colitis samples, 7 Crohn's disease samples and 3 control samples.

The data were analyzed using the Rosetta Resolver software. Statistical significance of the microarray data was determined by Student's unpaired *t* test. $p < 0.01$ and a fold change of greater or less than 1.5 were considered statistically significant. To correct for multiple hypothesis testing a q-value was calculated for each tested feature to estimate significance in terms of the false discovery rate (FDR) rather than the false positive rate. For every differential expression analysis the q-value was calculated and a FDR was calculated using the method proposed by Storey and colleagues.(185) A FDR of less than 5% was calculated for each of the presented analysis.

Hierarchical clustering analysis was undertaken using the Pearson correlation method. Genes that had a greater or less than 1.5 fold change in expression were included in the cluster analysis of the control, Crohn's disease and ulcerative colitis biopsies. Statistical significance

of clustering into specific disease groups was calculated by Chi squared analysis. Gene ontology was analyzed using Ingenuity software (Ingenuity Systems, Mountain View, CA). The Mann-Whitney U test was used to analyze the real time PCR data. $p < 0.05$ was considered significant.

3.4 GENOTYPING METHODS

3.4.1 Recruitment of Patients and Healthy Controls

The definitions of ulcerative colitis and Crohn's disease were based on the classifications as described by Lennard-Jones.(211) Patients were recruited from the gastroenterology units of The Western General Hospital, Edinburgh, Royal Infirmary, Edinburgh and St. John's Hospital, Livingston (all in Lothian region serving a catchment population of approximately 500 000 patients) between 1997-2005. Fourteen patients with ulcerative colitis and 7 patients with Crohn's disease were recruited to both the micoarray study as well as in the genetic studies. Patients were recruited by Ian Arnott, Jack Satsangi, Colin Noble, Janice Fennell and Linda Smith. Control patients were actively recruited from healthy volunteers and also obtained from anonymous Scottish-Caucasian blood donors from the Lothian region. The demographic details and numbers used are described in detail in the subsequent chapters. All patients and controls were coded anonymously prior to entry to database as described below. In each of these patients, 10 mls of whole blood were obtained for DNA extraction.

3.4.2 Phenotypic data acquisition and Storage

Phenotypic data were collected by Ian Arnott, Colin Noble, Hazel Drummond, Janice Fennell and Gwo-Tzer Ho. Phenotypic classification were based on the Vienna classification (115) and data were collected by patient questionnaire, interview and case-note review and comprised of demographics, date of onset and diagnosis, disease location, disease behaviour,

progression, extra-intestinal manifestations, surgical operations, smoking history, joint symptoms, family history and ethnicity. All original phenotypic data was stored in a *Microsoft Access Database*™, which was independently maintained by a database manager (Hazel Drummond). The main statistical analyses were performed using the *Minitab* v10

3.4.3 Ethics Approval

Written consent was obtained from each patient and the Medicine and Oncology Subcommittee of the Lothian Local Research Ethics Committee approved the study protocol: LREC 2000/4/192.

3.4.4 DNA extraction and storage

Genomic DNA was extracted by Norman Anderson in Edinburgh from peripheral venous blood by a modified salting-out technique and resuspended in 1xTE (10mM Tris (pH 8.0), 1mM EDTA (pH 8.0) at a final concentration of 100 ng/ml. The optical density of the resultant DNA was measured using a Gene Quant Pro instrument to estimate the actual DNA concentration. Using a standard protocol, 10 mls of whole blood were added into a 50ml conical tube together with 40 mls of red cell lysis buffer, and mixed and resuspended for 5 minutes. This was then centrifuged at 804 x g for 10 minutes. Following this, the supernatant was removed, the pellet resuspended with 10 mls of RCLB and re-centrifuged at 804 x g for a further 5 minutes. The pellet was then further resuspended in 3 mls of nuclear lysis buffer and sodium dodecyl sulphate. One ml of 6M sodium chloride and 3mls of chloroform were

subsequently added leading to emulsification of DNA (after vigorous mixing). This was then centrifuged at 804 x g for 20 minutes. The contents of the tube separated into 3 layers with the DNA in the middle layer. This layer was then removed and added into ethanol to allow precipitation of DNA. This DNA pellet was then removed and washed in 10mls of 70% ethanol and then dried in room air for 5 minutes. The precipitated DNA was transferred into 0.5 mls of TE buffer and stored at 4°C until it dissolved. Genomic DNA was diluted to 100ng/μl for PCR analysis.

3.4.5 PCR reaction: TaqMan® technology

The genotyping in this thesis was performed using TaqMan® by Angela Fawkes and Stuart Bayliss at the Wellcome Genetics Core, Edinburgh. This PCR-based assay uses laser scanning technology which excites fluorescent dyes present in the specially designed TaqMan® probes. The system includes a built-in thermal cycler, a laser to induce fluorescence, CCD (charge-coupled device) detector, real-time sequence detection software, and TaqMan® reagents for the fluorogenic 5' nuclease assay. The cycle-by-cycle detection of the increase in the amount of PCR product is quantified in real time as the special probes, "reporter dye", fluoresces when the "quencher" is removed from the fragment during the PCR extension cycle. The prepared genomic DNA was laid out in 96-well plates prior to genotyping by the Wellcome Genetics Core, Western General Hospital, Edinburgh. (Table 3.7) In each plate, negative control wells are used to check the quality of reaction. Thermal cycling conditions were two initial holds (50° C for 2 min and 95° C for 10 min) followed by a 40-cycle two-step program (95° C for 15 sec and 60° C for 1 min).

Table 3.7: TaqMan reaction volumes

Reaction Component	Assay by Design 40X mix
Abgene Absolute QPCR ROX mix	2.5µl
Genotyping Assay Mix	0.125µl
dH_2O	2.375µl

3.4.6 Allelic discrimination in TaqMan

Allelic discrimination was scored independently by 2 blinded-laboratory scientists (Angela Fawkes and Stuart Bayliss, Wellcome Genetics Core, Edinburgh). As quality control, a random selection of DNAs was re-genotyped and compared with original data. An example of the allelic discrimination plot was as shown in Figure 3.2.

Figure 3.2: TaqMan® allelic discrimination plot.

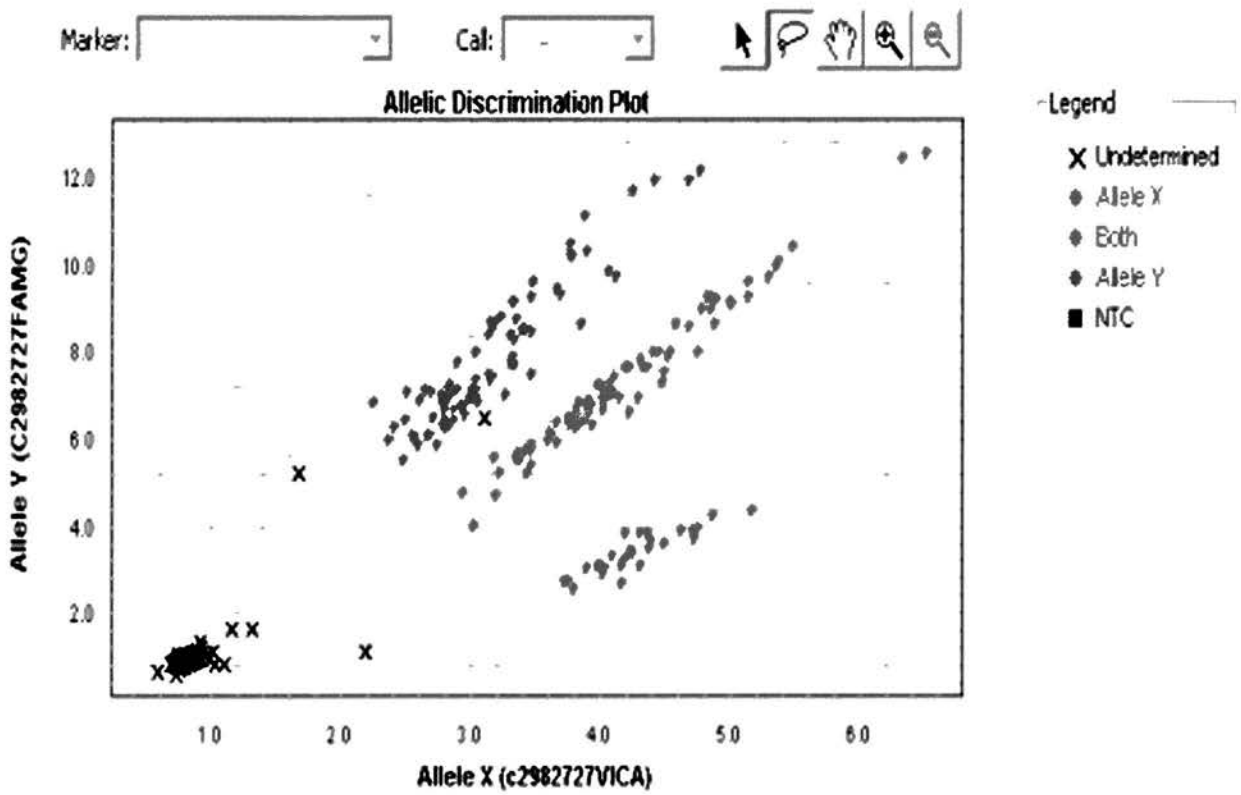


Figure 3.2: An example of the readout from the TaqMan® allelic discrimination PCR-based assay showing the results of the alleles.

3.4.7 Genetic Data Analysis

Each SNP was analyzed for association with inflammatory bowel disease overall, Crohn's disease, ulcerative colitis and disease phenotype. Each of the variants studied was shown to be in Hardy-Weinberg equilibrium in the control population. Genotype-phenotype associations were analyzed by chi squared test, using *Minitab* statistical software package. Linkage

disequilibrium was assessed using cocaphase software (Rosland Franklin centre for genomics research, www.hgmp.mrc.ac.uk) and results were presented as pairwise D′ values, a measure between 0 and 1 where 0 is equal to no correlation and 1 shows complete correlation. Haplotype analysis was investigated using Haploview (http://www.broad.mit.edu/mpg/haploview/). Evidence for epistasis was investigated by comparing allelic frequencies of the individual variants by chi squared analysis.

To identify significant independent variables associated with genotype, multiple logistic regression analysis was carried out using *Minitab* statistical software package. The population attributable risk percentage (PAR%) was defined as the excess rate of disease in individuals with a mutation compared with those without. This was estimated by the method of Schlesselman (215) which estimates PAR% as equal to attributable risk as a function of the prevalence in the exposed population, divided by incidence of IBD in the population. To calculate this, the prevalence of IBD was estimated at 100/100,000 and the frequency of all alleles in the control population was assumed to reflect that of the general population.

CHAPTER 4

CHARACTERIZATION OF INTESTINAL GENE EXPRESSION PROFILES IN CROHN'S DISEASE BY GENOME-WIDE MICROARRAY ANALYSIS

SUMMARY

Introduction and Aims: Genome-wide microarray expression analysis creates a comprehensive picture of gene expression at the cellular level. The aim of this study was to investigate differential intestinal gene expression in patients with Crohn's disease (CD) and controls with sub-analysis of confirmed CD susceptibility genes, associated pathways and cell lineages.

Methods: 53 CD and 31 control subjects-23 normal and 8 inflamed non-inflammatory bowel disease patients were studied. Paired endoscopic biopsies were taken from 5 specific anatomical locations for RNA extraction and histology. 41058 expression sequence tags were analyzed using the *Agilent* platform.

Results: When all the CD biopsies were compared to the controls, 259 sequences were upregulated and 87 sequences were downregulated. Upregulated genes in CD included SAA1 (FC +7.5, $p = 1.47 \times 10^{-41}$) and REGL (FC +7.3, $p = 2.3 \times 10^{-16}$). Downregulated genes included genes involved in cellular detoxification-SLC14A2 (FC -2.49, p =0.00002), CA2 (FC -2.4, $p = 8.4 \times 10^{-10}$) and CA1 (FC -2.3, $p = 7.5 \times 10^{-6}$). In the CD terminal ileal biopsies diubiquitin (FC+11.3, $p < 1 \times 10^{-45}$), MMP3 (FC+7.4, $p = 1.3 \times 10^{-11}$), IRTA1 (FC-11.4, $p = 4.7 \times 10^{-12}$) and CCL23 (FC-7.1, $p = 1.6 \times 10^{-10}$) were differentially expressed compared to controls. In the colon SAA1 (FC+6.3, $p = 5.3 \times 10^{-8}$) was upregulated and TSLP (FC-2.3, $p = 2.7 \times 10^{-6}$) was downregulated comparing non-inflamed CD and control biopsies, and the colonic inflammatory CD signature was characterised by downregulated organic solute carriers-SLC38A4, SLC26A2 and OST alpha. Analysis of the IL-23 pathway revealed IL-23A, JAK2 and STAT3 were upregulated in the CD group compared to controls and in the inflamed compared to non-inflamed CD biopsies. Modest changes in expression were also observed in a number of the autophagy genes.

Conclusion: Expression of a number of key inflammatory molecules and pathways were significantly dysregulated in CD emphasising their role in disease pathogenesis and potential for translation to therapeutic targets.

4.1 INTRODUCTION

Current evidence suggests that the inflammatory bowel diseases, Crohn's disease and ulcerative colitis are complex non-Mendelian polygenic disorders with important gene-environmental interactions.(34)

In the past 2 years, a number of genome-wide association studies (GWAS) in populations of European descent and a subsequent meta-analysis have identified 32 confirmed Crohn's disease susceptibility genes/loci.(151) These include innate immune genes that are specific to Crohn's disease; NOD2 (originally described in 2001)(42;43) and the autophagy genes ATG16L1 and IRGM,(57) clearly indicating that defects in the intracellular processing of bacteria constitute a central feature in the pathogenesis of Crohn's disease. The discovery that germline variants of IL23R were protective in Crohn's disease coincided with murine experiments detailing the contribution of IL23 (rather than IL12 with which it shares the p40 subunit) to Th17 driven chronic intestinal inflammation.(145;216) The meta-analysis and subsequent studies in ulcerative colitis have demonstrated that 3 other IL23 pathway genes (IL12B, JAK2 and STAT3) are all inflammatory bowel disease susceptibility genes.(151)

At present there are no large scale intestinal genome-wide expression studies in Crohn's disease. There is now an immediate need to explore in detail the function and expression of the novel genetic associations.

A number of microarray studies have now been carried out in immune cell subsets to try to understand differences in gene expression during activation and inflammation. Genome wide expression from a compendium of six immune cell types has allowed investigators to identify a collection of immune response *in silico* genes that have specific expression in immune cells.(217) These genes have allowed investigators to differentiate signaling pathways in immune cell subsets and to charcterize the inflammatory response of genes known to play a role in immune response and genes of unknown function.

The aim of the present study was to use microarray expression analysis to describe the transcriptional profiles in the colon and the terminal ileum in patients with Crohn's disease and controls. In additional to this hypothesis-free scanning, expression of germ line variants identified by GWAS and cell specific lineage analysis were also investigated.

4.2 MATERIALS AND METHODS

53 patients with Crohn's disease and 31 control patients who did not have inflammatory bowel disease and who were undergoing colonoscopy were recruited (Table 4.1). Patient recruitment and data collection are detailed in the methods section (Chapter 3.2.1). Combined analysis of all the inflammatory bowel disease biopsies was also undertaken, adding in biopsies from patients with ulcerative colitis (Table 5.1).

4.2.1 Biopsy Collection

Paired biopsies were taken from each anatomical location. One biopsy was sent for histological examination and the other was snap frozen in liquid nitrogen for RNA extraction (Chapter 3.1.2-3). The number of paired biopsies from each anatomical location are shown in Table 4.2.

Table 4.1: The demographics of the Crohn's disease and control patients.

	Crohn's disease	**Controls**
Number of patients	53	31
Male/ Female	26/27	11/20
Median age at diagnosis (years)	28.6	43 at time of endoscopy
Median duration of follow up (years)	8.1	
Disease Group		
New Diagnosis (1)	7 (13%)	
Quiescent disease (2)	30 (57%)	
Active disease (3)	16 (30%)	
Vienna Classification of disease location at diagnosis		
Ileal disease (L1)	7 (13%)	
Colonic disease (L2)	33 (62%)	
Ileo-colonic disease (L3)	11 (21%)	
Vienna Classification of disease location at endoscopy		
Ileal disease (L1)	6 (11%)	
Colonic disease (L2)	28 (53%)	
Ileo-colonic disease (L3)	19 (36%)	
Vienna Classification of disease behavoiur at diagnosis		
Inflammatory (B1)	40 (75%)	
Stricturing (B2)	3 (6%)	
Penetrating (B3)	7 (13%)	
Vienna Classification of disease behavoiur at endoscopy		
Inflammatory (B1)	32 (60%)	
Stricturing (B2)	8 (16%)	
Penetrating (B3)	12 (23%)	
Surgery*	20 (38%)	
Current Smoker	11 (21%)	
Family history of IBD	12 (23%)	
Extra-articular symptoms	13 (25%)	
5 ASA Therapy	21 (40%)	
Corticosteroid therapy	4 (8%)	
Immunosuppressant therapy (AZA, 6MP, MTX, MMF)	13 (25%)	

* Includes patients who were operated on for luminal complications of Crohn's disease. Full phenotypic data were available on 94% of patients at the time of diagnosis and 100% of patients at the time of endoscopy.

Table 4.2: The location, number and inflammation status of biopsies from Crohn's disease patients and controls

	Crohn's disease		Controls	
Total number of paired biopsies	106		76	
Removed from analysis	7		3	
	Inflamed	**Non-inflamed**	**Inflamed**	**Non-inflamed**
Terminal Ileum	10	6	1	5
Ascending colon	12	8	3	14
Descending colon	14	16	6	17
Sigmoid colon biopsies.	16	17	8	19

4.2.2 Microarray Analysis

Detailed methods are provided the methods section (Chapter 3.2.4-8). Total RNA was extracted from each biopsy and one 1µg of total RNA was amplified using the low RNA input fluorescent linear amplification protocol.

The samples were hybridized for 18 hours at 60°C with constant rotation. Microarray slides were scanned using the *Agilent* G2505B model. Expression signals were calculated using the *Agilent* feature extraction software.

4.2.3 Real Time PCR

Real time PCR analysis was undertaken in 4 genes-IL-8, SAA1, DEFA5 & 6 using RNA from 15 Crohn's disease and 6 control terminal ileal biopsies. SAA1 and IL-8 were selected as they are robust markers of inflammation and showed significant changes in the microarray data set. DEFA5 and A6 were selected as we had a particular interest in their expression in the terminal ileum. Detailed methods are described in the methods (Chapter 3.3.1). TaqMan primers and probes were manufactured in house (*Genentech, Inc*) and are shown in Table 4.3.

Table 4.3: TaqMan primers used for real- time PCR.

Gene		Sequence
SAA1	forward	agcgatgccagagagaata
	Reverse	ggaagtgattggggtctttg
	Taq	ctttggccatggtgcggagg
IL8	forward	actcccagtcttgtcattgc
	Reverse	caagtttcaaccagcaagaa
	Taq	tgtgttggtagtgctgtgttgaattacgg,
DEFA5	forward	gctacccgtgagtccctct
	Reverse	tcttgcactgctttggtttc
	Taq	tgtgtgaaatcagtggccgcct
DEFA6	forward	agagctttgggctcaacaag
	Reverse	atgacagtgcaggtcccata
	Taq	cacttgccattgcagaaggtcctg

4.2.4 Data Analysis

Statistical significance of the microarray data was determined by Student's unpaired *t* test. $p < 0.01$ and a fold change of greater or less than 1.5 were considered statistically significant. To correct for multiple hypothesis testing a q-value was calculated for each tested feature to estimate significance in terms of the false discovery rate (FDR) rather than the false positive rate. For every differential expression analysis the q-value was calculated and a FDR was calculated using the method proposed by Storey and colleagues.(185) A FDR of less than 5% was calculated for each of the presented analysis. Gene ontology was analyzed using Ingenuity software.

The Mann-Whitney U test was used to analyze the real time PCR data. $p < 0.05$ was considered significant. Gene ontology was analyzed using Ingenuity software.

Hierarchical clustering analysis using a collection of immune response *in silico* genes from a compendium of six immune cell types was undertaken.(217) Hierarchical clustering analysis was also undertaken using a set of 14 epithelial cell cytokines-

CXCL1, CXCL2 CXCL5, CXCL9, CXCL10, CXCL11, CCL2, CCL4, CCL7, CCL20, IL-8, IL-12A, IL-23A and MDK.(218-220)

4.3 RESULTS

4.3.1 Gene Expression in Crohn's Disease and Controls

When 99 Crohn's disease biopsies were compared to 73 control biopsies, 259 sequences were upregulated and 87 sequences were downregulated. The 20 most upregulated and downregulated sequences are shown in (Table 4.4) and the complete table is Supplementary Table 1 on the Appendix 1 CD. Notably upregulated genes in the Crohn's disease biopsies included the acute phase proteins serum amyloid A1, SAA1, (FC +7.5, $p = 1.47 \times 10^{-41}$), the regenerating C-type lectin family member REGL, (FC +7.3, $p = 2.3 \times 10^{-16}$), the acute phase proteins S100A9, (FC +4.4, $p = 2.4 \times 10^{-22}$) and S100A8, (FC +4.0, $p = 3.5 \times 10^{-18}$). IL-8 a robust marker of mucosal inflammation was the sixth most upregulated gene (FC +3.6, $p = 5.6 \times 10^{-19}$) and a protein that interacts with TNFAIP3 (A20) TNIP3 was also increased (FC +3.8, 4.2×10^{-6}). Among the most downregulated genes were genes involved in cellular detoxification- SLC14A2, (FC -2.49, p =0.00002), carbonic anhydrase 2, (FC -2.4, $p = 8.4 \times 10^{-10}$) and carbonic anhydrase 1, (FC -2.3, $p = 7.5 \times 10^{-6}$).

Table 4.4: Expression changes in the 40 most dysregulated sequences comparing all the Crohn's disease biopsies to control biopsies.

Gene Name	Sequence Code	Fold Change	P-value
SAA1	A_24_P335092	7.47764	1.47E-41
REGL	A_23_P108546	7.26194	2.25E-16
S100A9	A_23_P23048	4.37037	2.37E-22
S100A8	A_23_P434809	4.00494	3.48E-18
TNIP3	A_23_P386478	3.83902	4.17E-06
IL8	A_32_P87013	3.60471	5.60E-15
IF	A_24_P92472	3.52236	6.18E-13
KCND3	A_32_P140268	3.38296	2.37E-18
CLECSF12	A_24_P235988	3.28581	6.65E-10
chromosome 10 open reading frame 81	A_23_P23980	3.19877	6.76E-13
regenerating islet-derived 3 gamma	A_32_P65628	3.15414	6.74E-07
TFEC	A_32_P184394	3.10908	5.09E-12
IGSF6	A_23_P106629	2.99982	1.28E-12
A_32_P90385	A_32_P90385	2.99826	5.10E-14
GW112	A_23_P2789	2.84338	5.99E-19
MGC27165	A_24_P315941	2.63927	2.00E-06
MMP3	A_23_P161698	2.62676	3.91E-10
KLK12	A_23_P500010	2.60484	2.40E-11
TZFP	A_23_P131024	2.57661	1.09E-08
REG4	A_24_P58673	2.56805	1.39E-12
CLECSF9	A_24_P78531	2.56703	2.91E-06
AATK	A_23_P10559	-1.71307	1.55E-13
ECT2	A_24_P366033	-1.72142	2.52E-06
SLC26A2	A_23_P250951	-1.72588	0.00003
XRRA1	A_23_P370162	-1.73618	6.37E-07
RPS28	A_24_P40010	-1.77456	3.88E-17
ISL1	A_23_P81529	-1.79059	5.87E-07
MGC29643	A_23_P419696	-1.79074	4.43E-09
AQP8	A_23_P26522	-1.79996	0.00004
FLJ25770	A_24_P401185	-1.85266	0.00031
IL1R2	A_23_P79398	-1.86364	5.39E-11
ANKRD17	A_24_P220771	-1.87438	1.63E-06
A_32_P191066	A_32_P191066	-1.89029	1.62E-06
FLJ12572	A_24_P65121	-1.90062	0.00052
LOC339881	A_24_P846810	-1.94299	1.28E-10
NKD1	A_24_P304881	-2.10407	1.48E-17
CA1	A_23_P168916	-2.26411	7.46E-06
PRAC	A_23_P15619	-2.42192	4.16E-11
CA2	A_23_P8913	-2.44317	8.36E-10
LOC389023	A_32_P86578	-2.48381	2.18E-28
SLC14A2	A_24_P136471	-2.49075	0.00002

4.3.2 Gene Expression in the Terminal Ileum

When all of the CD terminal ileal (TI) biopsies were compared to control TI biopsies 1035 sequences had a fold change of greater than 1.5 and 1014 sequences had a fold change of less than 1.5 ($p<0.01$). The 20 most upregulated and downregulated sequences are shown in (Table 4.5) and the complete table is Supplementary Table 2.

Upregulated genes in the Crohn's disease biopsies included diubiquitin (UBD) which is involved in protein degradation in eukaryotic cells, (FC +11.3, $p <1x10^{-45}$), MMP3, (FC +7.4, $p =1.3x10^{-11}$), IL-8, (FC +4.9, $p =2.3x10^{-8}$), trefoil factor 1 (TFF1) which acts in the GI tract to maintain the mucosal surface barrier, (FC +4.3, $p =1.3x10^{-7}$) and the cytokeratin keratin 5β, (FC +4.2, $p =0.005$) (Table 4.6). Downregulated genes included immune associated genes IRTA1- a novel surface B-cell receptor, (FC -11.1, $p =4.7x10^{-12}$), CCL23, (FC -7.1, $p =1.6x10^{-10}$), CXCR4, (FC -6.0, $p =8.2x10^{-18}$), and genes involved in cholesterol metabolism APOC3, (FC -8.2, $p =7.0x10^{-8}$) and APOA1, (FC -6.9, $p =0.0031$).

Table 4.5: Expression changes in the 40 most dysregulated sequences comparing Crohn's disease and control terminal ileal biopsies.

Gene Name	Sequence Code	Fold Change	P-value
UBD	A_23_P81898	11.30144	<10E-45
TIMD4	A_32_P69616	10.1666	1.21E-08
FLJ25393	A_24_P305993	9.5289	0.00061
FLJ27099	A_32_P200144	9.09174	2.98E-32
SOX14	A_32_P183652	8.89445	2.28E-14
BX108833	A_24_P460405	8.28796	4.11E-08
HK2	A_32_P175739	7.76749	5.97E-19
MMP3	A_23_P161698	7.42185	1.29E-11
RP11-653A5.1	A_32_P84237	7.41899	2.75E-07
TEX12	A_23_P150362	7.1498	7.01E-09
III	A_32_P157391	7.06942	2.35E-10
S100P	A_23_P58266	6.37114	3.88E-28
C1orf34	A_23_P160214	6.28438	4.85E-18
Sprn	A_24_P930415	5.92864	0.00002
FOLH1	A_23_P47616	5.8971	1.55E-20
LOC92552	A_23_P361744	5.3328	6.97E-06
EYA2	A_23_P500421	5.32674	0.00091
CEACAM3	A_23_P130515	5.29423	1.44E-06
C14orf81	A_24_P323298	5.29169	8.53E-08
MGC27165	A_23_P259763	5.26036	0.00012
HEBP1	A_23_P117082	-4.43185	3.24E-09
ARHGAP24	A_32_P72067	-4.48983	7.05E-10
LOC375180	A_32_P49764	-4.52586	3.49E-12
SUSD2	A_23_P314101	-4.683	6.76E-06
AGXT2	A_24_P63468	-4.84685	0.00011
CYFIP2	A_24_P465879	-4.89606	0.0001
FNBP1	A_24_P899020	-4.92271	7.33E-19
SLC28A2	A_23_P48816	-5.02797	1.05E-15
OTTHUMP011522	A_23_P45821	-5.03803	2.68E-08
PAX8	A_23_P324916	-5.80485	0.00263
CXCR4	A_23_P102000	-6.01856	8.18E-18
CGGA17790	A_32_P203728	-6.46865	0.00016
APOA1	A_23_P203191	-6.8623	0.00305
C6orf32	A_24_P941359	-7.02034	1.09E-13
NPPC	A_24_P174353	-7.14015	0.00395
CCL23	A_24_P133905	-7.14857	1.62E-10
APOC3	A_23_P203183	-8.17858	7.02E-08
IRTA1	A_23_P115200	-9.62094	6.82E-06
MGC27169	A_23_P407695	-9.89477	4.82E-08
IRTA1	A_23_P115201	-11.4284	4.72E-12

Table 4.6: Expression changes in genes of interest in biopsies from the terminal ileum.

Gene	Sequence Code/ Genbank cluster code	All Crohn's disease (CD) samples (16) v controls (6). Fold change (FC)	p value	CD Non-inflamed (6) v non- inflamed controls (6) (FC)	p value	CD Inflamed (10) v non-inflamed (6) (FC)	p value
APOA1	A_23_P203191	-6.86	0.0031	-1.032	0.91	-12.22	0.00003
APOC3	A_23_P203183	-8.18	7.02×10^{-8}	+1.36	0.10	-12.36	9.70×10^{-14}
CD28	A_23_P91015	-3.76	1.77×10^{-17}	-4.52	1.32×10^{-22}	+1.30	0.12
CXCL13	A_23_P121695	-3.18	2.1×10^{-6}	-32.00	0.00002	+8.1	0.010
CXCR4	A_23_P102000	-6.02	8.2×10^{-18}	-2.1	5.23×10^{-10}	+1.73	0.0033
DefA5	A_23_P112086	-1.16	0.034	-1.07	0.41	-1.14	0.22
DefA6	A_23_P363711	-1.085	0.11	-1.11	0.34	+1.04	0.70
IL-8	A_32_P87013	+4.85	2.30×10^{-8}	+1.63	0.0017	+16.9	1.26×10^{-13}
IRTA1	A_23_P115201	-11.43	4.72×10^{-12}	-1.57	0.0001	-2.93	0.0032
TFF3	A_23_P257296	+2.40	$<10^{-45}$	+2.0	1.47×10^{-16}	+1.72	6.1×10^{-22}
UBD	A_23_P81898	+11.3	$<10^{-45}$	+8.48	1.32×10^{-34}	+2.50	0.00009

Fold changes and p values are shown in a number of different genes in three different experiments. The number of biopsies analyzed in each experiment is shown in brackets. Candidate genes were included in this table if significant consistent changes in expression were observed across more than one experiment.

4.3.3 Colonic Gene Expression Analysis

To minimize the effect of differential gene expression related to the anatomical location of the biopsy, sigmoid colon biopsies were used for analysis. To also remove the acute inflammatory expression signature non- inflamed Crohn's disease biopsies (n=17) were compared to non- inflamed control biopsies (n=18) (Supplementary Table 3, Appendix 1 CD). SAA1 remained the most upregulated gene, (FC +6.3, p $=5.3x10^{-8}$) and in total 279 sequences were upregulated. 349 sequences were downregulated and the most downregulated genes included MMP1, (FC -3.6, p $=2.4x10^{-15}$), CXCL13, (FC -2.7, p =0.005) and TSLP-thymic stromal lymphoprotein; (FC -2.3, p $=2.7x10^{-6}$) (Table 4.7).

When the acute inflammatory signal was examined in the sigmoid colon and 16 inflamed Crohn's disease biopsies were compared to 17 non-inflamed Crohn's disease biopsies, 279 sequences were upregulated and 148 sequences were down regulated (Supplementary Table 4). Upregulated genes in the inflamed biopsies included OLFM4- an anti-apoptotic molecule that inhibits the capsase cascade and also binds to GRIM19, (FC +6.2, p $=2.9x10^{-14}$) and TNIP3. Downregulated genes included organic solute carriers SLC38A4, (FC -2.7, p= 0.005), SLC26A2, (FC -2.5, p =0.00001) and OST alpha, (FC -2.5, p =0.008).

Table 4.7: Expression changes in genes of interest in biopsies from Crohn's disease (CD) patients and controls.

Gene	Sequence code	All CD (99) v controls (73) All CD Fold change (FC)	p value	Inflamed (16) v non-inflamed (17) CD sigmoid (FC)	p value	Inflamed CD sigmoid (16) v inflamed control sigmoid (9) (FC)	p value	Non-inflamed CD sigmoid (17) v non-inflamed control (18) sigmoid Fold change	p value
IFNG	A_23_P151294	+2.1	$2.3x10^{-9}$	+2.0	0.0080	+1.29	0.50	+1.37	0.18
IL-8	A_32_P87013	+7.5	$1.5x10^{-41}$	+2.5	0.0088	+3.35	0.0030	+1.06	0.59
MMP3	A_23_P52761	+2.63	$3.9x10^{-10}$	+2.3	0.0029	+7.6	$3.14x10^{-10}$	-1.50	0.015.
SAA1	A_24_P335092	+7.5	$1.5x10^{-41}$	+3.6	$5.6 x10^{-15}$	+8.1	$1.4x10^{-7}$	+6.3	$5.3x10^{-8}$
TNF	A_23_P376488	-1.079	0.0031	+1.26	0.0044	-1.13	0.15	-1.10	0.13
TNIP3	A_23_P386478	+3.84	$4.2x10^{-6}$	+3.63	$2.9x10^{-10}$	+4.41	0.00008	-1.27	0.27.
TSLP	A_23_P121987	-1.52	0.00021	-1.19	0.34	-1.42	0.49	-2.34	$2.7x10^{-6}$

Fold changes and p values are shown in a number of different genes in four different experiments. The number of biopsies analyzed in each experiment is shown in brackets. Novel genes identified by analysis of the microarray data set and genes with an established role in the pathogenesis of inflammatory bowel disease were investigated.

4.3.4 Analysis of Gene Expression Changes in Inflammatory Bowel Disease Biopsies and Controls

To look for overall changes in gene expression, all 99 Crohn's disease biopsies and control biopsies were analyzed alongside 129 ulcerative colitis biopsies, giving a total of 228 inflammatory bowel disease biopsies which were then compared to the 73 control biopsies. 154 sequences had a fold change of greater than 1.5 and 37 sequences had a fold change of less than 1.5 ($0.01>p>10^{-45}$) (Supplementary Table 5). The most upregulated gene was serum amyloid A1 SAA1, (FC +7.15, $p = < 1x10^{-45}$) followed by TNFAIP3, (FC +6.0, $p = 6.7x10^{-17}$), which is involved in the TNF signaling pathway and then the S100 proteins S100A8, (FC +3.9, $p = 1.4x10^{-34}$) and S100A9, (FC +3.7, $p = 1.7x10^{-26}$). A smaller number of genes were downregualted including SLC14A2, (FC -1.9, p = 0.0024) and the fold changes observed were smaller.

4.3.5 Expression of Genes implicated by GWAS Meta-analysis

Expression of susceptibility genes identified by GWAS meta- analysis by *Barrett and colleagues* (151) were investigated along with further detailed analysis of the IL-23 and autophagy pathways (Table 4.8). Upregulated genes with modest increases in expression in the Crohn's disease biopsies compared to the controls included NOD2/ CAD15, (FC +1.23, p =0.000243), PTGER4- prostaglandin E receptor 4, (FC +1.1, p =0.00010) and NKX2.3, a 3 exon homeobox gene, (FC +1.37, p =0.001). There was no expression probe on the *Agilent* microarray chip that represented IGRM and no

differences were observed between disease groups when expression of TNFSF15, PTPN22, ICOSLG, ITLN1, ZNF365, LRRK2 and PTPN2 were examined.

When inflamed and non-inflamed Crohn's disease sigmoid colon biopsies were compared MST1-Macrophage stimulatory protein; FC -1.58, p =0.0037 was downregulated in the inflamed biopsies.

Table 4.8: Expression of genes identified by *Barrett and colleagues* (151) as being associated with Crohn's disease

Entrez Gene			All CD (99) v controls (73)		Inflamed CD sigmoid (16) v non-inflamed CD sigmoid (17)	
ID	Symbol	Agilent ID	Fold Change	p value	Fold Change	p value
717	JAK2	A_23_P123608	+1.90	9.43E-07	+1.58	0.000031
5054	ATG16L1	A_32_P113508	-1.16	1.96E-05	+1.06	0.549
593	IL-23A/ p19	A_23_P425197	+2.32	0.000099	+2.11	0.000031
734	PTGER4	A_23_P435394	+1.11	0.000104	-1.04	0.55
4127	NOD2	A_23_P420863	+1.23	0.000243	+1.24	0.1092
774	STAT3	A_24_P116805	+2.23	0.000353	+1.66	0.0002
59296	NKX2-3	A_23_P52425	+1.37	0.000994	-1.17	0.456
4901	CDKAL1	A_23_P44781	-1.1	0.00964	-1.14	0.0919
4103	ORMDL3	A_23_P38190	+1.13	0.0140	+1.07	0.656
6946	C11orf30	A_23_P380839	+1.1	0.0156	-1.22	0.0077
966	TNFSF15	A_23_P94754	+1.08	0.0447	+1.09	0.5281
6191	PTPN22	A_23_P201181	+1.07	0.107	+1.03	0.6849
235	CCR6	A_24_P234921	+1.21	0.144	+1.84	0.0566
3308	ICOSLG	A_23_P317667	+1.1	0.161	-1.10	0.857
5600	ITLN1	A_23_P95790	-1.1	0.162	-1.02	0.905
2891	ZNF365	A_23_P86610	+1.17	0.244	-1.22	0.423
20892	LRRK2	A_23_P128447	+1.25	0.413	+1.37	0.135
771	PTPN2	A_23_P309701	-1.04	0.483	+1.07	0.545
485	MST1	A_24_P148796	-1.04	0.709	-1.58	0.0036
49233	IL-23R	A_23_P7560	-1.02	0.823	+1.05	0.4271

For each experiment the fold change and p values have been calculated. The number of biopsies analyzed in each experiment are shown in brackets.

4.3.6 The IL-23 Pathway

When Crohn's disease samples were compared to controls IL-23A/p19, (FC +2.32, p =0.000099), JAK2, (FC +1.90, p =$9.4x10^{-7}$), STAT3, (FC +2.23, p =0.0004) and INFγ, (FC +2.31, p =0.0019) were significantly upregulated in the Crohn's disease biopsies (Figure 4.1). When inflamed Crohn's disease biopsies were compared to non-inflamed Crohn's disease biopsies IL-23A/p19, (FC +2.11, p =0.000031), JAK2, (FC +1.90, p =0.00003), STAT3, (FC +1.66, p =0.0002) and INFγ, (FC +2.33, p <0.0001) had increased expression in the inflamed biopsies. No significant changes were observed in IL-23R expression.

Figure 4.1: Expression analysis of the IL-23/ Th17 pathway in Crohn's disease and controls

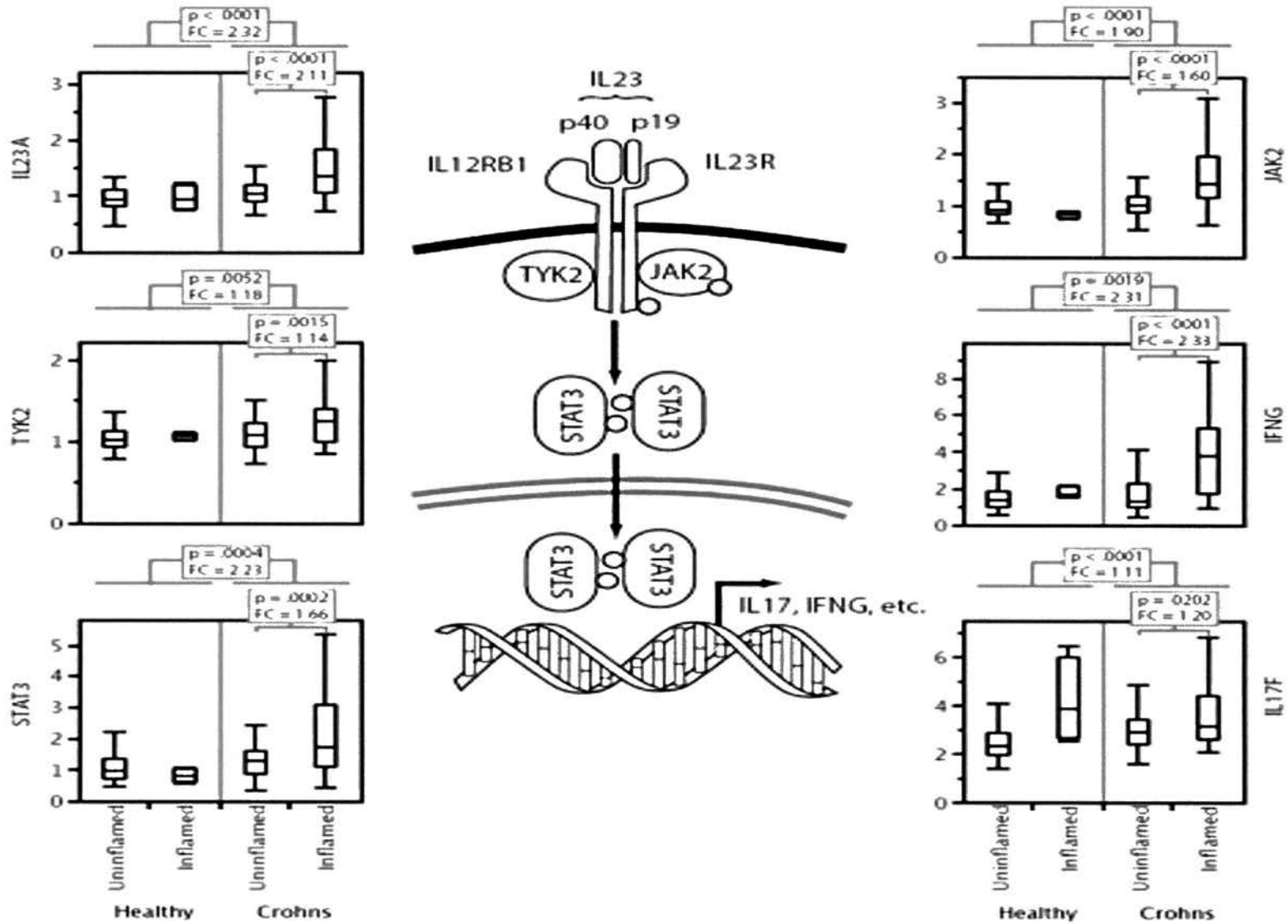


Figure 4.1 The IL-23 pathway (Alex Abbas, Genentech, personal communication) is illustrated along with gene expression of constituent molecules in Crohn's disease and control biopsies separated by inflammation status. Gene expression is shown as box-whisker plots. The boxes are 25th to the 75th centile. The IL-23 pathway is upregulated in Crohn's disease biopsies compared to controls and in inflamed Crohn's disease biopsies compared to non-inflamed Crohn's disease biopsies.

4.3.7 Autophagy Pathway

ATG16LI and 19 other genes and key regulators of the autophagy pathway were analyzed (Figure 4.2). ATG16LI was modestly downregulated in the CD biopsies regardless of inflammation status compared to controls; FC -1.16, $p=1.96x10^{-5}$ as was (ATG4D; FC -1.14, p=0.0007) and (ATG3; FC -1.06, p=0.0052). (ATG12; FC +1.1, p=0.041), (ATG16L2; FC +1.1, p=0.045) and (LC3B; FC +1.18, p=0.0003) were marginally upregulated in the CD biopsies compared to the controls.

Figure 4.2: Expression analysis of the autophagy pathway in Crohn's disease and controls

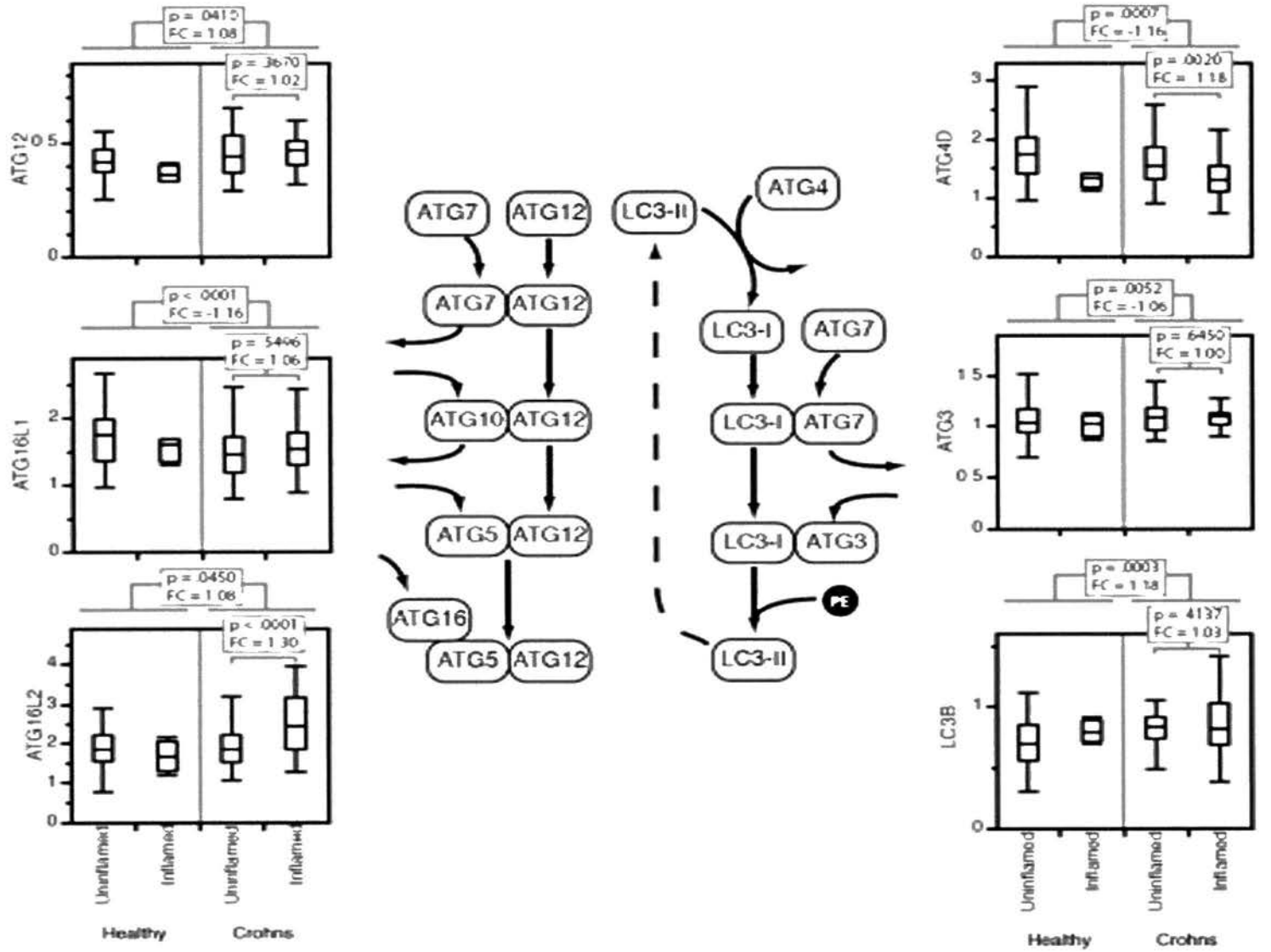


Figure 4.2 The autophagy pathway is illustrated (Alex Abbas, Genentech, personal communication). Gene expression is shown as box- whisker plots. PE – Phosphatidylethanolamine, a lipid which covalently attaches to ATG8/LC3 and mediates its attachment to autophagic membranes.

4.3.8 Unsupervised Hierarchical Clustering Analysis

When all of the Crohn's disease (n= 99) and control biopsies (n= 73) were clustered together using unsupervised hierarchical clustering analysis, no separation of the biopsies by either disease status or by the degree of inflammation was observed. When the anatomical location that the biopsies were taken from was considered, 18 terminal ileal biopsies clustered together (6 control and 12 Crohn's disease) ($X^2 = 10.2$, $p < 0.001$).

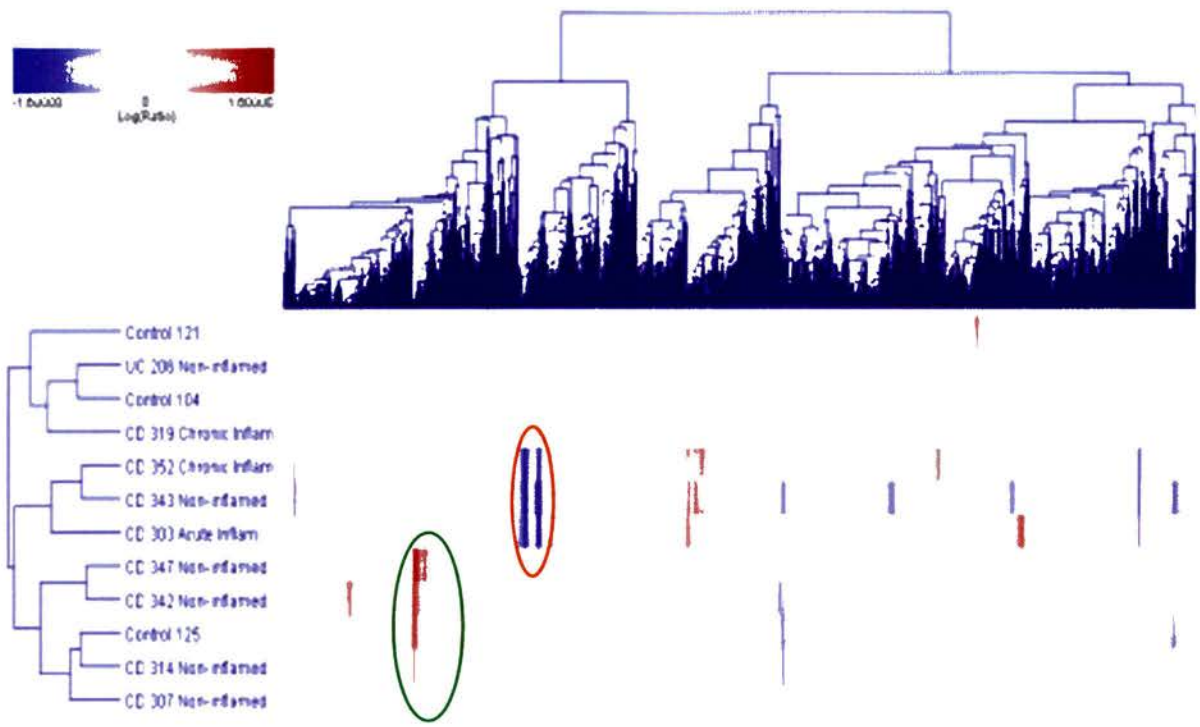
Unsupervised clustering analysis of the terminal ileal biopsies initially was confounded by the sex of patients, however when a degree of supervision was introduced and only terminal ileal biopsies from female patients and controls were clustered, clustering by disease status was observed (Figure 4.3).

Gene ontology of the 593 downregulated sequences grouped by biological process revealed a preponderance of genes associated with carboxylitic acid metabolic processes- (39 of a total of 464 genes classified by the ontology software to this biological group; OR 3.4, $p = 7 \times 10^{-13}$), organic acid metabolic processes (38/464; OR 3.1 $p = 1 \times 10^{-12}$) and lipid metabolic processes (46/620; OR 3.0, $p = 6.6 \times 10^{-12}$). When the downregulated sequences were grouped by biological function genes grouped under solute/ cation transporter activity (11/50; OR 10.3, p

$=6.9x10^{-15}$), electrochemical potential- driven transporter activity (23/188; OR 5.16, p $=2.7x10^{-14}$) and solute/ sodium transporter activity (10/46; OR 10.1, p $=2.4x10^{-13}$) were disproportionately downregulated. When these groups of genes were combined to encompass all genes involved in transporter activity, there was a significant over representation of this group in the downregulated genes (64/1138; OR 2.3, p $=3.6x10^{-9}$).

367 sequences were upregulated in a subset of the Crohn's disease samples compared to the controls. Ontology of these genes grouping by biological processes showed that genes that grouped into structural molecule activity (22/603; OR 2.62, p $=4.5x10^{-5}$) and extracellular matrix structural constituents (6/87; OR 5.5, p =0.0003) were overrepresented. When the genes were grouped by biological function upregulated genes grouped into sequence specific DNA binding (11/430; OR 2.28, p =0.007) and transcription factor activity (20/810; OR 1.7, p =0.043).

Figure 4.3: Hierarchical clustering of terminal ileal biopsies from females with Crohn's disease and controls.



Terminal ileal biopsies from 8 patients with Crohn's disease, three healthy controls with normal terminal ileal pathology and one patient with ulcerative colitis who had normal terminal ileal pathology were clustered. The Crohn's disease, ulcerative colitis and control patients are annotated with the inflammation status of the biopsy. The degree of upregulation measured in red and downregulation measured in blue can be quantified using the logarithmic key. Two areas appeared to be driving this separation and these have been highlighted in orange-downregulated and green-upregulated.

4.3.9 Hierarchical Clustering by Specific Probe Subsets: *Immune Response in Silico (IRIS)* Probes

Using the previously defined *IRIS* probes to detect differential expression, Crohn's disease and control biopsies from the ascending and descending colon were compared.(217) Using the B cell, monocyte and T cell probes we were able to observe separation of the biopsies into Crohn's disease and control biopsies by unsupervised clustering- B cell probes ($p = 0.0006$, OR 2.74), the monocyte probes ($p < 0.0001$ OR 5.22) and the T cell probes ($p = 0.0047$ OR 2.4) using Chi squared analysis. In the monocyte cluster 2 genes CXCL1 and MMP1 were markedly differentially regulated in the Crohn's disease biopsies and controls. No terminal ileal clustering was observed for any of the examined probes.

4.3.10 Hierarchical Clustering by Epithelial Cell Markers

Unsupervised clustering analysis using a panel of 14 epithelial cell cytokines showed clear separation between colonic biopsies from Crohn's disease patients and controls $p < 0.00001$ (Figure 4.4). When terminal ileal biopsies were considered this separation was not observed ($p = 0.052$).

Figure 4.4: Sigmoid colon Crohn's Disease and control biopsies clustered by Epithelial Cell Markers.

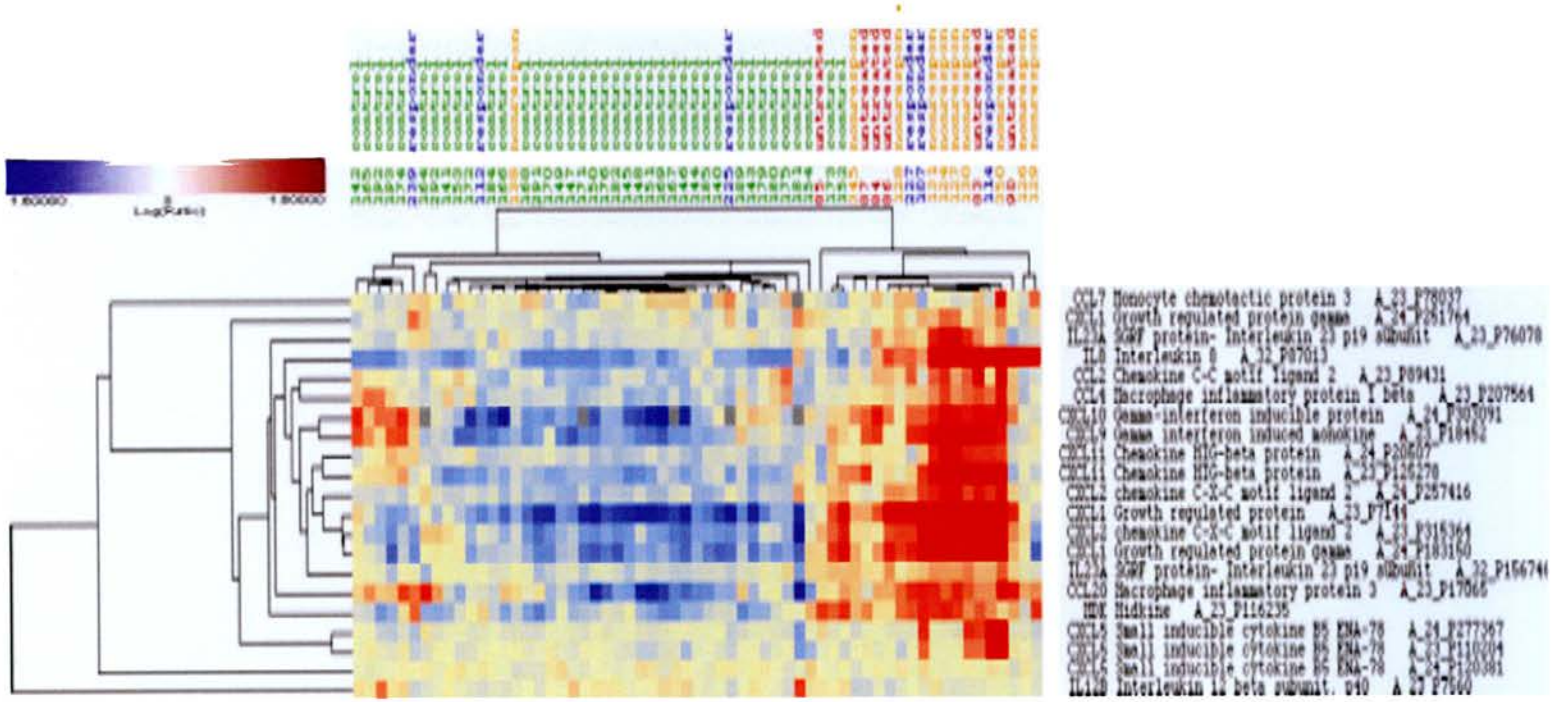


Figure 4.4 The colonic biopsies are annotated along the top of the figure (green-controls, blue-non-inflamed Crohn's disease, orange-inflamed Crohn's disease, red-untreated inflamed Crohn's disease). On the right of the figure the epithelial cell cytokines are annotated. The degree of upregulation measured in red and downregulation measured in blue can be quantified using the logarithmic key.

4.3.11 Real Time PCR Confirmation of Microarray Results

In line with the histological classification of the biopsies, and the microarray results significantly higher IL-8 levels were observed in the Crohn's disease terminal ileal biopsies compared to the control terminal ileal biopsies ($p= 0.0045$) and in the inflamed Crohn's disease terminal ileal biopsies compared to the non-inflamed Crohn's disease terminal ileal biopsies ($p= 0.0046$)(Table 4.9). Trends were also observed towards SAA1 being more highly expressed in the Crohn's disease biopsies compared to the controls and in the inflamed compared to the non-inflamed ileal Crohn's disease biopsies. No difference in DEFA5 & 6 expression was observed in the Crohn's disease terminal ileal biopsies compared to the control terminal ileal biopsies ($p= 0.73$ and $p= 0.97$ respectively), nor when the inflamed Crohn's disease terminal ileal biopsies were compared to non-inflamed Crohn's disease terminal ileal biopsies ($p= 0.39$ and $p= 0.69$ respectively).

Table 4.9: Real time PCR expression in terminal ileal biopsies of patients with Crohn's disease and controls

Genes	Median Relative Expression in control TI biopsies (6)	Median Relative Expression in CD TI biopsies (15)	Median Relative Expression in non- inflamed CD TI biopsies (7)	Median Relative Expression v inflamed CD TI biopsies (8)	Median Relative Expression in inflamed (8) v non- inflamed (7) CD TI biopsies
IL-8	8.5	65.7 (0.0045)	20.1 (0.054)	307 (0.0037)	307 v 20.1 (0.0046)
Def A5	1.26	0.70 (0.73)	0.51 (0.43)	0.96 (0.95)	0.98 v 0.51 (0.39)
Def A6	0.87	1.07 (0.97)	1.1 (0.74)	1.04 (0.85)	1.04 v 1.1 (0.69)
SAA1	1.7	3.52 (0.20)	2.0 (0.52)	20.7 (0.14)	20.7 v 2.0 (0.18)

The relative expression of each gene is shown in disease groups along with the p value in brackets. The p values are calculated compared to the control group for each gene analyzed. The number of biopsy samples used in each analysis is also shown in brackets.

4.4 DISCUSSION

In this accurately phenotyped data set, strict separation of the biopsies by anatomical location and inflammatory status has allowed us to minimize background noise and real time PCR confirmation has increased the confidence associated with the interpretation of the data. For the large number of novel Crohn's disease susceptibility genes from GWAS where little data are presently available, we have been able to investigate expression profiles in the human colon and terminal ileum.

When all the Crohn's biopsies were compared to controls serum amyloid A1 (SAA1) was the most upregulated gene, followed by regenerating C-type lectin family member (REGL) and the S100A8 and A9 genes. SAA1 is a HLA- associated apolipoprotein acute phase reactant and levels can be elevated in inflammation, trauma and neoplasia. Its transcription is induced by the pro- inflammatory cytokines IL-2, IL-6, TNFα and bacterial LPS, and it is the major factor responsible for the development of secondary AA amyloidosis in chronic immune mediated diseases such as Rheumatoid arthritis or Crohn's disease.(221) Ray et al have speculated that persistently elevated levels of SAA1 contribute to reactive AA amyloidosis.(222) In Crohn's disease reactive AA amyloidosis is rare and a much more attractive role for SAA1 would be as a marker of disease activity, severity, and potentially because of its induction by TNFα a predictor of response to anti- TNF therapy.

Given current concerns with respect to the reproducibility of microarray expression data it is firstly reassuring that our results are consistent with the findings from a previous microarray

study in Crohn's disease patients where increased expression of the S100 and the REG gene families was also observed.(167) Furthermore, in parallel with the results of Costello et al we observed a number of sequences representing novel proteins that were differentially expressed and using ontology analysis we were able characterize genes into functions related to Crohn's disease pathogenesis.(196)

This is the first study where genome wide expression has been investigated in unpooled terminal ileal endoscopic biopsies from Crohn's disease patients and controls. The most upregulated gene in the Crohn's disease compared to the control terminal ileal biopsies was diubiquitin or ubiquitin-like protein FAT10. The family of ubiquitin-like proteins function as part of the ubiquitin proteasome system which is a crucial pathway for protein degradation in eukaryotic cells.(223) The gene is located at the major histocompatibility complex locus in on chromosome 6 (224), an established Crohn's disease susceptibility loci and its expression has been observed to be increased in 90% of hepatocellular carcinomas and in 80% of colon cancers.(225)

Diubiquitin is a downstream target of p53 and in p53-defective cells its expression is increased resulting in chromosomal instability.(226;227) Overall in this data set diubiquitin was upregulated when all Crohn's disease biopsies were compared controls by a fold change of 1.5 and interestingly diubiquitin expression in hepatocellular cancer and colon cancer correlates with increased expression of IFN-gamma and TNFα suggesting a mechanism for carcinogenesis in this pro-inflammatory environment.(228)

In the terminal ileal biopsies there was no difference in expression of the alpha defensins 5 and 6 (DEFA5&6) in the Crohn's disease patients and controls regardless of the degree of inflammation in the biopsies. These results were confirmed by real time PCR and are contrary to previous data where reduced DEFA5&6 expression was observed in the terminal ileum of Crohn's disease patients regardless of the degree of inflammation.(66)

More recently, Simms et al also showed that expression of DEFA5&6 was downregulated in terminal ileal Crohn's disease biopsies.(229) However, this downregulation was inflammation specific, probably reflecting a loss of the epithelial layer and a reduction of epithelial and Paneth cells as a consequence of persistent inflammation. Previously we have shown that the increase in colonic expression of DEFA5&6 in ulcerative colitis patients is largely mediated by Paneth cell metaplasia and that in the colon unregulated Paneth cell differentiation, and the consequent increase in DEFA5&6 expression, may perpetuate mucosal inflammation.(230)

An interesting change in expression in the colonic Crohn's disease biopsies reflecting the traditional Th1 and novel Th17 paradigm in Crohn's disease was the downregulation of thymic stromal lymphopoietin (TSLP) in non-inflamed colonic Crohn's disease samples compared to non-inflamed controls. TSLP is a cytokine that mediates its effect through dendritic cells to promote the Th2 differentiation of $CD4^+$ T cells.(231) Moreover, mice with an intestinal epithelial cell (IEC) deletion of intrinsic IκB kinase, have reduced TSLP expression and as a consequence have a poor Th2 immune response resulting in an inability to eradicate infection.(232) These mice also develop severe intestinal inflammation as a result of dendritic cell derived Th1 and Th17 pathway activation and it is intriguing to speculate that in

the non-inflamed human Crohn's disease colon decreased levels of TSLP may perpetuate the subsequent persistent and excessive inflammation.

When the colonic analysis was compared to our previous expression studies in ulcerative colitis there was a 23% homology between the differentially regulated genes in the respective Crohn's disease and ulcerative colitis analysis compared to controls.(230) The colonic inflammatory expression signature observed in the Crohn's disease biopsies was also similar to that observed in the ulcerative colitis biopsies and one of the most differentially regulated genes in both of the data sets was serum amyloid A1 (SAA1).

The identification of IL-23R as a Crohn's disease susceptibility gene has focused investigation towards the distinct Th17 lineage.(54) We observed that expression of a number of components of this pro-inflammatory pathway- IL-23A, STAT3, JAK2 and IFNγ were increased in Crohn's disease compared to controls and that this change was driven by active as opposed to quiescent disease. These convincing genetic and expression data emphasize the importance of this pro-inflammatory pathway in the pathogenesis of Crohn's disease. Multiple therapeutic targets have been identified in this pathway and clinical trials of a monoclonal antibody against the p40 subunit of IL-12/23 have produced promising early clinical data.(59)

The discovery of ATG16L1 as a Crohn's disease specific susceptibility gene has strongly implicated the autophagy pathway in the pathogenesis of Crohn's disease. Autophagy is a highly conserved cellular process where the cell digests part of its own cytoplasm and it functions as a normal physiological response to remove toxic material or intracellular bacteria

from the cell. The pathway has also been implicated in the pathogenesis of neurodegenerative diseases such as Alzheimer's and Parkinson's disease.(52)

Recent data have linked the innate immune response and autophagy via Toll-like-receptor (TLR) engagement.(233) TLR induced phagosomes within macrophages triggered ATG5 and ATG7 mediated acidification and enhanced killing of the ingested organisms. These interactions between the innate immune system and the autophagy pathway have provoked investigators to speculate about specific interaction between NOD2/CARD15 and autophagy and this is an area of active investigation.

NKX2.3, a novel IBD candidate gene was upregulated in Crohn's disease and it has been observed that NKX2.3 deficient mice have smaller spleens and Peyers patches, and that in these animals the T and B cells are disorganized.(234) Furthermore, NKX2.3 homozygote mice lack mucosal adhesion cell molecule-1 (MAdCAM-1), probably as a result of a lack of NKX2.3 mediated MAdCAM-1 transcription. MAdCAM-1 mediates lymphocyte homing and extravasation during the inflammatory response and this pathway has clinical implications with regards therapeutic intervention and monitoring response to treatment.(235)

Clustering analysis has allowed us to differentiate between biopsies from Crohn's disease patients and controls and the observed separation was driven by a cluster of downregulated genes involved in the normal homeostasis of the terminal ileum-organic acid and lipid metabolic processes, and solute/cation transporter activity. The cluster of upregulated genes had a preponderance of genes that grouped into structural molecule activity function. The

differing expression signature observed in the terminal ileal biopsies appeared to be primarily inflammation driven, rather than disease specific as the changes were less obvious in the non-inflamed analysis than when the inflamed and non-inflamed Crohn's disease biopsies were compared. These dysregulated probes could form the basis of a diagnostic expression chip to help diagnose ileal Crohn's disease and grade its severity.

An alternative method of clustering analysis was to cluster samples using a subset of genes related to cell lineage.(217) We have undertaken this analysis in our samples by separating by genes from key immune cell types and observing clustering of the colonic biopsies. From this we can clearly identify immune cell infiltration in the biopsies and characterize the most differentially expressed genes. These expression signatures can also be used to gain insight into genes of unknown function and provide a resource to investigate immune cell differentiation in health and in different immune mediated diseases.

A final area of interest was in the role of the intestinal epithelial cell in the inflammatory process. The fourteen intestinal epithelial cell markers we investigated showed good ability to segregate Crohn's disease patients and controls by clustering analysis with the majority of the chemokines being upregulated in the colonic Crohn's disease biopsies in an inflammation dependant manner. These results are consistent with previous data from Puleston and colleagues who observed a subset of chemokines- CXCLs 1-3 and CCL20 were upregulated in colonic inflammatory bowel disease along with their receptors in a coordinated intestinal epithelial cell inflammatory response.(236) The upregulation of these chemokines was

significantly more than known leukocyte chemokines emphasizing the central role of the intestinal epithelial cell in colonic inflammation.

Further studies carried out in human colonic inflammatory bowel disease biopsies, in human colonic cell lines and in human fetal intestinal xenografts have all confirmed the central role of the intestinal epithelial cell in mediating, coordinating and perpetuating the pathogenic inflammatory response observed in the colon in both Crohn's disease and ulcerative colitis.(218;237;238)

In conclusion this valuable data set has allowed us to gain novel insight into the pathogenesis of Crohn's disease at the mucosal level. A number of key regulators of intestinal inflammation were observed to be differentially expressed. Furthermore, the data add considerably to the recent genome wide association studies in providing complimentary human colonic and ileal expression data along with detailed analysis of the IL-23 and autophagy pathways. In depth analysis of these exciting new candidate genes along with intestinal epithelial cell specific analysis have generated a number of potential therapeutic targets worthy of further investigation.

CHAPTER 5

REGIONAL VARIATION IN GENE EXPRESSION IN THE HEALTHY COLON IS DYSREGULATED IN ULCERATIVE COLITIS

SUMMARY

Background and Aims: Microarray analysis allows a comprehensive picture of gene expression at the cellular level. The aim of this study was to investigate differential intestinal gene expression in patients with ulcerative colitis (UC) and controls.

Methods: 67 UC and 31 control subjects- 23 normal and 8 inflamed non-inflammatory bowel disease patients were studied. Paired endoscopic biopsies were taken from 5 specific anatomical locations for RNA extraction and histology. 41058 expression sequence tags were analyzed in 215 biopsies using the *Agilent* platform. Confirmation of results was undertaken by real time PCR and immunohistochemistry.

Results: In healthy control biopsies, cluster analysis showed differences in gene expression between the right and left colon. (χ^2=25.1, p<0.0001). Developmental genes HOXA13, ($p=2.3x10^{-16}$), HOXB13 ($p<1x10^{-45}$), GLI1 ($p=4.0x10^{-24}$), and GLI3 ($p=2.1x10^{-28}$) primarily drove this separation. When all UC biopsies and control biopsies were compared, 143 sequences had a fold change of >1.5 in the UC biopsies ($0.01>p>10^{-45}$) and 54 sequences had a fold change of <-1.5 ($0.01>p>10^{-20}$). Differentially upregulated in UC genes included SAA1 ($p<10^{-45}$) the alpha defensins, DEFA5&6 (p=0.00003 and $p=6.95x10^{-7}$ respectively), MMP3 ($p=5.6x10^{-10}$) and MMP7 ($p=2.3x10^{-7}$). Increased DEFA5&6 expression was further characterized to Paneth cell metaplasia by immunohistochemistry and *in-situ* hybridization. Sub-analysis of the IBD2 loci, and the ABC transporter genes revealed a number of differentially regulated genes in the UC biopsies.

Conclusions: These data implicate a number of novel gene families, as well as established candidate genes in the pathogenesis of UC, and may allow characterization of potential therapeutic targets.

5.1 INTRODUCTION

In recent years, the application of non-parametric linkage analyses and well-designed case-control association studies have led to the identification of a number of susceptibility genes or loci strongly associated with Crohn's disease and ulcerative colitis. Candidate genes and loci implicated in ulcerative colitis include the IBD2, and IBD5 loci, HLA complex and the MDR1 gene.(34) A recent genome wide association study identified a novel susceptibility locus ECM1 that was associated with ulcerative colitis (152) and further genome wide association studies carried out in patients of European ancestry have identified ulcerative colitis risk loci on chromosomes 1p36 and 12q15(153), and a SNP flanking IL-10, a strong candidate gene in ulcerative colitis pathogenesis.(239) These studies all highlight important pathways involved in disease pathogenesis and the emerging challenge now is to move from gene identification to functional understanding.

The aims of the current study were to use microarray gene expression analysis to investigate genome wide expression in endoscopic mucosal biopsies taken at colonoscopy from 67 patients with ulcerative colitis and 31 controls. This represents the largest cohort described to date and particular care has been taken to characterize disease phenotype as well as the anatomical location sampled. The data extend current understanding of gene expression in health, and are complementary to current studies of germline and somatic variation associated with ulcerative colitis.

5.2 METHODS

5.2.1 Patients and Controls

67 patients with ulcerative colitis and 31 control patients undergoing colonoscopy were recruited (Table 5.1) (Methods 3.2.1).

Table 5.1: Demographics of the ulcerative colitis patients

	Ulcerative colitis
Number of patients	67
Male/ Female	33/34
Median age at diagnosis (years)	37
Median duration of follow up (years)	7.8
Disease Group	
New Diagnosis	8
Quiescent disease	41
Active disease	18
Disease extent at time of Endoscopy	
Proctitis	15
L sided colitis	27
Extensive colitis	25
Current Smoker	6
Family history of IBD	5
5 ASA Therapy	40
Corticosteroid therapy	10
Immunosuppressant therapy (AZA, 6MP, MTX)	11

Eleven of the controls were male, 20 were female with a median age of 43 at the time of endoscopy. Six of the controls had normal colonoscopies for colon cancer screening, 9 controls had symptoms consistent with irritable bowel syndrome and had a normal colonoscopic investigation and 7 patients had a colonoscopy for another indication and histologically normal biopsies were obtained. Eight control patients had abnormal inflamed

colonic biopsies (1 pseudomembranous colitis, 1 diverticulitis, 1 amoebiasis, 2 microscopic colitis, 1 eosoinophilic infiltrate, 2 scattered lymphoid aggregates and a history of gastroenteritis). Written informed consent was obtained from all patients.

5.2.2 Biopsy Collection

Paired biopsies were taken from each anatomical location (Table 5.2).

Table 5.2: The location, number and inflammation status of biopsies in ulcerative colitis patients and controls

	Ulcerative colitis		**Controls**	
Total number of paired biopsies	139		76	
Removed from analysis	10		3	
	Inflamed	**Non- Inflamed**	**Inflamed**	**Non-Inflamed**
Terminal Ileum	0	4	1	5
Ascending colon	12	21	3	14
Descending colon	15	20	6	17
Sigmoid colon biopsies.	35	22	8	19

5.2.3 Microarray Analysis

Detailed methods are provided in the methods section (Methods 3.2.5-8). Total RNA was extracted from each biopsy using the micro total RNA isolation from animal tissues protocol, according to the manufacturer's instructions. One 1μg of total RNA was amplified using the low RNA input fluorescent linear amplification protocol. The samples were hybridized for 18 hours at 60°C with constant rotation. Microarray slides were scanned using the *Agilent* G2505B model. Expression signals were calculated using the *Agilent* feature extraction software (*version 7.5*).

5.2.4 Real Time PCR, I*n Situ* Hybridization for Defensin Alpha 5 and Immunohistochemistry for Rabbit Anti-Human Lysozyme and Rabbit Anti-Human Defensin Alpha 6

Methods are provided in the methods section 3.3.1-3.

5.2.5 Data Analysis

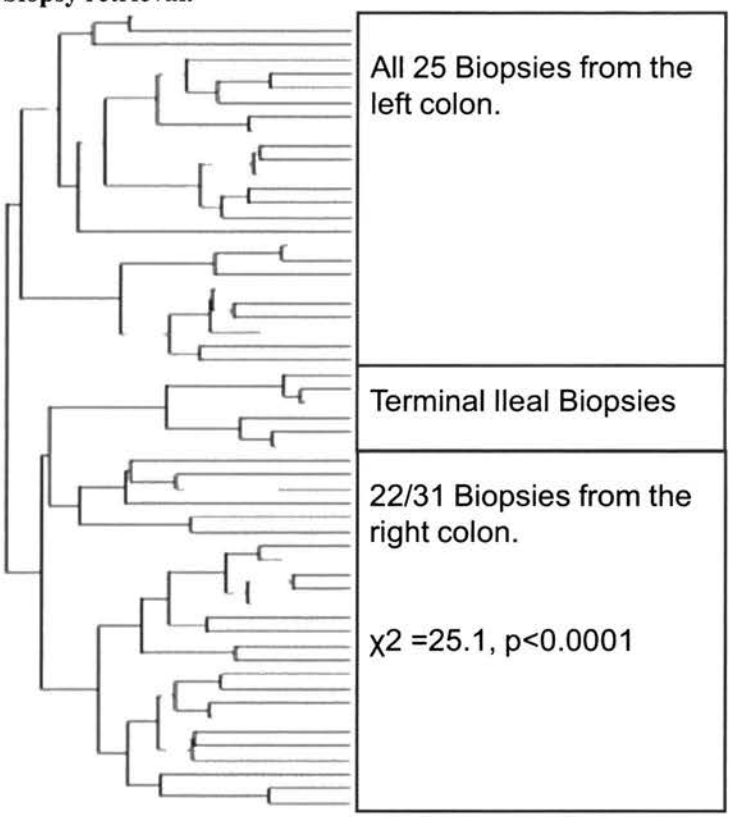
Microarray data were analyzed using the Rosetta Resolver software. Statistical significance of the microarray data was determined by Student's unpaired *t* test. $p < 0.01$ and a fold change of greater or less than 1.5 were considered statistically significant. To correct for multiple hypothesis testing a q-value was calculated for each tested feature to estimate significance in terms of the false discovery rate (FDR) rather than the false positive rate. From the q-values a FDR was calculated using the method proposed by Storey and colleagues.(185) A FDR of less than 5% was calculated for each of the presented analysis. Gene ontology was analyzed using Ingenuity software. The Mann-Whitney U test was used to analyze the real time PCR data. $p < 0.05$ was considered significant.

5.3 RESULTS

5.3.1 Influence of anatomical location on gene expression in the healthy colon and terminal ileum

Unsupervised hierarchical clustering analysis using probes that had a fold change of greater or less than 1.3 were used to interrogate 56 histologically normal biopsies from control patients. Clear separation by anatomical location was observed- on one side of the dendrogram 25/25 biopsies were from the left colon (descending colon or sigmoid colon) where as on the other side of the dendrogram 20/31 biopsies were from the ascending colon (χ^2 =25.1, p<0.0001) (Figure 5.1). Biopsies from individual patients did not cluster together. The genes driving the differential expression between the right and left colon that were causing the observed clustering were predominantly involved in the embryological development of the GI tract- Homeobox proteins HOXA13, Fold Change (FC) +4.93, p= $2.3x10^{-16}$), HOXB13 (FC+16.96, p <$1x10^{-45}$), Glioma-associated oncogenes GLI1 (FC+2.2, p= $4.0x10^{-24}$), and GLI3 (FC+2.3, p= $2.1x10^{-28}$) were all upregulated in the left colon. 61 sequences had a fold change of greater than 1.5 and 44 sequences had a fold change of less than 1.5 in the left colon in the control biopsies. In the ulcerative colitis biopsies when non- inflamed left and right colonic biopsies were compared 26 sequences had greater than a 1.5 fold increase in expression and 21 sequences had less than a -1.5 fold decrease in the left colon.

Figure 5.1: Unsupervised hierarchical clustering of 56 healthy control biopsies showing the location of biopsy retrieval.



Histologically normal biopsies from control patients were analyzed by unsupervised hierarchical clustering. Separation of the biopsies by anatomical location was observed and this separation was predominantly driven by genes involved in the embryological development of the GI tract.

When gene ontology was compared between these two analyses alternative splicing genes were the most differentially regulated genes in both of the analysis ($p= 1.7x10^{-41}$ and $p= 7.9x10^{-19}$ for controls and ulcerative colitis respectively). In the control analysis the next 3 most differentially regulated gene groups were nuclear protein function ($p= 4.3x10^{-29}$), metal binding ($p= 4.7x10^{-25}$) and membrane function ($p= 4.2x10^{-20}$). In ulcerative colitis oxidoreductase function ($p= 1.4x10^{-13}$) was the 2nd most differentially expressed group

followed by mitochondrial function ($p= 4.7x10^{-25}$) and catalytic activity ($p= 2.7x10^{-11}$). There was a 34.5% overlap between the differentially expressed ontology groups in the control and ulcerative colitis analysis.

5.3.2 Analysis of expression in ulcerative colitis and control biopsies

Using unsupervised hierarchical clustering we were unable to differentiate between biopsies from ulcerative colitis patients and controls patients. In addition no clustering based on the inflammation status of the biopsies was observed. When all of the ulcerative colitis biopsies (129) and control biopsies (73) were compared, 143 sequence probes had a fold change of greater than 1.5 in the UC biopsies ($0.01 > p > 10^{-45}$) and 54 sequences had a fold change of less than 1.5 ($0.01 > p > 10^{-20}$). The 20 most upregulated and downregulated sequences are shown in (Table 5.3) and the complete table is Supplementary Table 6 on the Appendix 1 CD. Serum amyloid A1 (SAA1) was the most upregulated gene FC+8.18, $p < 10^{-45}$. Other notably upregulated genes were the S100 calcium-binding proteins A8 (FC+3.50, $p = 2.3x10^{-17}$), S100A9 (FC+3.06, $p = 4.1x10^{-13}$), the alpha defensins, alpha 5 (DEFA5) (FC+3.25, p = 0.00003), alpha 6 (DEFA6) (FC+2.18, $p = 6.95x10^{-7}$) and the matrix metalloproteinases MMP3 (FC+2.17, $p = 5.6x10^{-10}$) and MMP7 (FC+2.29, $p = 2.3x10^{-7}$). The differential gene expression of a number of candidate genes of interest is shown in Table 5.4.

Gene ontology analysis involving the genes differentially expressed between the ulcerative colitis and control biopsies showed a preponderance of differentially expressed genes were involved in immune response (48 genes out of a total of 679 genes classified under immune

response, p= $2.1x10^{-9}$, OR 2.61, CI 1.85-3.56) and response to wounding (30 genes out of a total of 359 genes classified under response to wounding, p= $6.42 x10^{-9}$, OR 3.14, CI 2.09-4.53) when biological systems were considered.

5.3.3 Analysis of expression in sigmoid colon biopsies in patients with quiescent ulcerative colitis and non-inflamed control biopsies

To compare expression in biopsies without an acute inflammatory signal and to remove the effect of anatomical variation, 22 biopsies from the sigmoid colon with no histological evidence of inflammation from patients with ulcerative colitis were compared to 18 histologically normal control sigmoid colon biopsies. 102 sequences had a fold change greater than 1.5 ($0.01 > p > 4.77x10^{-13}$) and 84 sequences had a fold change of less than 1.5 ($0.01 > p > 1.8x10^{-21}$). The 20 most upregulated and downregulated sequences are shown in (Table 5.5) and the complete table is Supplementary Table 7. Upregulated genes in the ulcerative colitis biopsies included defensin beta 14 (FC+2.11, p = 0.00002) and SAA1 (FC+2.01, p = 0.00024). Interesting genes that were down regulated included HLA-DRB1 (FC-3.0, p = 0.0010) and TSLP (FC-2.73, $p = 2.7x10^{-10}$) (Table 5.4).

Table 5.3: Expression changes in the 40 most dysregulated sequences comparing ulcerative colitis biopsies to controls.

Gene Name(s)	Sequence Code	Fold Change	P-value
SAA1	A_24_P335092	8.17928	<10E-45
TNIP3	A_23_P386478	8.02411	1.13E-17
S100A8	A_23_P434809	3.49944	2.31E-17
DEFA5	A_23_P112086	3.24658	0.00003
S100A9	A_23_P23048	3.0603	4.17E-13
TZFP	A_23_P131024	2.68388	3.03E-10
UNG2	A_23_P92860	2.60581	2.50E-32
KCND3	A_32_P140268	2.56611	2.39E-09
MMP7	A_23_P52761	2.29453	2.24E-07
chromosome 10 open reading frame 81	A_23_P23980	2.25666	3.36E-07
C10orf81	A_24_P286951	2.24629	8.06E-12
regenerating islet-derived 3 gamma	A_32_P65628	2.22868	0.00041
DEFA6	A_24_P363711	2.1836	6.95E-07
MMP3	A_23_P161698	2.17171	5.62E-10
MGC27165	A_23_P259763	2.15704	5.92E-06
S100P	A_23_P58266	2.10221	3.53E-15
EP400	A_24_P298939	2.08571	4.10E-12
GW112	A_23_P2789	2.04692	5.47E-09
IL8	A_32_P87013	2.04605	4.24E-11
C14orf81	A_24_P608268	2.04054	0.00006
LOC92935	A_23_P108492	-1.66981	3.27E-06
GBA2	A_24_P341187	-1.67013	4.57E-15
RNF150	A_24_P350589	-1.67698	0.00004
LOC346329	A_24_P127159	-1.68116	2.51E-14
CST	A_23_P120863	-1.68211	3.82E-18
CLDN8	A_23_P427014	-1.70615	0.00532
H6PD	A_24_P626850	-1.70685	6.14E-10
MSTP9	A_23_P340376	-1.71901	1.34E-08
A_24_P792748	A_24_P792748	-1.73936	5.83E-11
SESTD1	A_23_P367610	-1.75701	7.07E-07
membrane-bound transcription protein 1	A_24_P378368	-1.75966	4.43E-12
NKD1	A_24_P304881	-1.76319	2.06E-11
LOC92552	A_23_P361744	-1.76501	0.00398
LOC339881	A_24_P846810	-1.80788	1.76E-07
ECT2	A_24_P366033	-1.83638	1.54E-08
PRAC	A_23_P15619	-1.84271	8.02E-06
LGALS2	A_23_P120902	-1.89554	6.36E-12
POLK	A_24_P919863	-2.14028	5.75E-11
LOC389023	A_32_P86578	-2.60671	7.74E-27
XIST	A_24_P500584	-3.28307	0.0064

Table 5.4: Expression changes in genes of interest.

Genes Analyzed	All UC (129 biopsies) v controls (73 biopsies) Fold change (FC)	p value	Non-inflamed UC sigmoid (22) v non-inflamed control sigmoid (18) (FC)	P value	Inflamed UC sigmoid (35) v inflamed control sigmoid (8) (FC)	p value	Inflamed UC sigmoid (35) v non-inflamed sigmoid UC (22) (FC)	p value
SAA1	+8.18	$<10^{-45}$	+2.0	0.00024	+17.5	$2.9x10^{-21}$	+16.5	$<10^{-45}$
Def alpha 5	+3.25	0.00003	+1.02	0.89	+7.27	$6.3x10^{-30}$	+8.44	$<10^{-45}$
Def alpha 6	+2.18	$6.95x10^{-7}$	-1.09	0.34	+4.41	$9.7x10^{-9}$	+6.72	$4.16x10^{-19}$
S100A8	+3.50	$2.3x10^{-17}$	+1.21	0.19	+9.75	$2.4x10^{-24}$	+6.84	$1.16x10^{-19}$
S100A9	+3.06	$4.1x10^{-13}$	+1.05	0.16	+7.53	$6.4x10^{-12}$	+7.11	$1.96x10^{-32}$
MMP3	+2.17	$5.6x10^{-10}$	-1.55	0.0088	+11.0	$1.22x10^{-37}$	+8.15	$2.32x10^{-35}$
MMP7	+2.29	$2.3x10^{-7}$	+1.16	0.080	+7.31	$4.9x10^{-24}$	+5.53	$1.01x10^{-23}$
IL8	+2.05	$4.2x10^{-11}$	+1.10	0.26	+6.36	$9.27x10^{-17}$	+7.24	$8.42x10^{-19}$
TLR4	+1.34	$4.5x10^{-7}$	+1.15	0.18	+1.50	0.0044	+1.54	0.00073
TNIP3	+8.02	$1.1x10^{-17}$	-1.30	0.20	+7.53	$2.93x10^{-13}$	+10.5	$1x10^{-38}$
CCL20	+1.30	0.00011	+1.25	0.020	+1.79	0.00002	+2.36	$4.68x10^{-11}$
ABCB1	-1.32	0.00091	+1.10	0.40	-1.82	$5.6x10^{-6}$	-1.92	$9.0x10^{-10}$
HLA-DRB1	+1.03	0.88	-3.0	0.0010	+3.30	0.033	+2.67	0.0011
TSLP	-1.12	0.31	-2.73	$2.7x10^{-10}$	-1.15	0.61	+1.23	0.092

Fold changes and p values are shown in a number of different genes in four different experiments. The number of biopsies analysed in each experiment is shown in brackets. Genes of interest were included in this table if significant consistent changes in expression were observed across more than one experiment. Gene annotation- TLR4- toll like receptor 4, TNIP3- TNFAIP3 (A20) -interacting protein 3, CCL20- chemokine (C-C motif) ligand 20, ABCB1- ABC transporter, HLA-DRB1- HLA class II histocompatibility antigen, DRB1, TSLP- thymic stromal lymphopoietin isoform 1.

Table 5.5: Expression changes in the 40 most dysregulated sequences comparing quiescent sigmoid colon ulcerative colitis biopsies to non-inflamed control sigmoid colon biopsies.

Gene Name(s)	Sequence Code	Fold Change	P-value
D2S448	A_32_P41327	6.8545	1.12E-06
LOC285189	A_32_P197825	5.07286	3.27E-06
AL359654	A_32_P865343	4.79967	0.00017
BG221366	A_32_P118811	4.68596	0.00005
BC034913	A_32_P217128	4.55265	0.00174
ATPase, H+ transporting, lysosomal V0 subunit	A_24_P923415	3.59124	5.26E-06
BX108833	A_24_P460405	3.41807	0.00424
TIMD4	A_32_P69616	3.30842	0.00008
RP11-653A5.1	A_32_P84237	3.06817	0.00859
LOC92552	A_23_P361744	3.02489	0.00044
FLJ31842	A_32_P19539	2.92736	7.86E-08
FLJ25393	A_24_P305993	2.78732	0.00654
TTTY15	A_24_P348861	2.58047	0.00006
QKI	A_23_P81760	2.50397	0.00311
GSCL	A_23_P109382	2.27819	0.00001
DEFB14	A_24_P931533	2.10838	0.00002
LOC286207	A_24_P229638	2.07887	0.0023
AQP8	A_23_P26522	2.02749	0.00003
SAA1	A_24_P335092	2.01116	0.00024
THC1975338	A_24_P681218	1.95708	1.67E-10
EDG1	A_23_P404481	-1.73072	0.00107
CD86	A_24_P131589	-1.73515	0.00075
PLAA	A_24_P78161	-1.76733	0.00116
THC1923453	A_32_P35668	-1.80382	0.00301
LOC284058	A_32_P210106	-1.81585	0.00456
TRA1	A_24_P150361	-1.86218	1.80E-21
BX119852	A_24_P640617	-1.88796	0.00818
chromosome 8 open reading frame 4	A_23_P253350	-1.89434	5.38E-06
LOC63929	A_24_P915710	-1.94095	0.00044
FGF12	A_24_P334300	-1.9687	0.00392
LOC339903	A_23_P80551	-2.08124	2.41E-11
CLIC6	A_23_P385067	-2.14902	1.32E-08
LOC285331	A_23_P396981	-2.18323	0.00106
LOC92935	A_23_P108492	-2.26078	0.0035
CTSZ	A_23_P40240	-2.27699	0.00056
MGC10814	A_23_P79032	-2.30755	5.22E-13
TTID	A_23_P110764	-2.38686	0.00002
TSLP	A_23_P121987	-2.73367	2.73E-10
HLA-DRB1	A_24_P169013	-3.0072	0.00101
XIST	A_24_P500584	-10.3502	0.00012

5.3.4 Inflamed versus non-inflamed ulcerative colitis sigmoid colon biopsies

When expression signals were compared between 35 histologically inflamed and 22 non-inflamed sigmoid colon ulcerative colitis biopsies, 700 sequences had a fold change of greater than 1.5 ($0.01 > p > 1x10^{-45}$) and 518 sequences ($0.01 > p > 1x10^{-45}$) had a fold change of less than 1.5 in the inflamed biopsies. The 20 most upregulated and downregulated sequences are shown in (Table 5.6) and the complete table is Supplementary Table 8. Notably upregulated genes included SAA1 (FC+16.5, $p = <10^{-45}$), TNFAIP3 interacting protein 3 (TNIP3) (FC+10.5, $p = 1x10^{-38}$), DEFA5 (FC+8.44, $p = <10^{-45}$), DEFA6 (FC+6.72, $p = 4.16x10^{-19}$) and regenerating islet-derived 3 gamma (REG3γ) (FC+6.99, $p = <10^{-45}$).

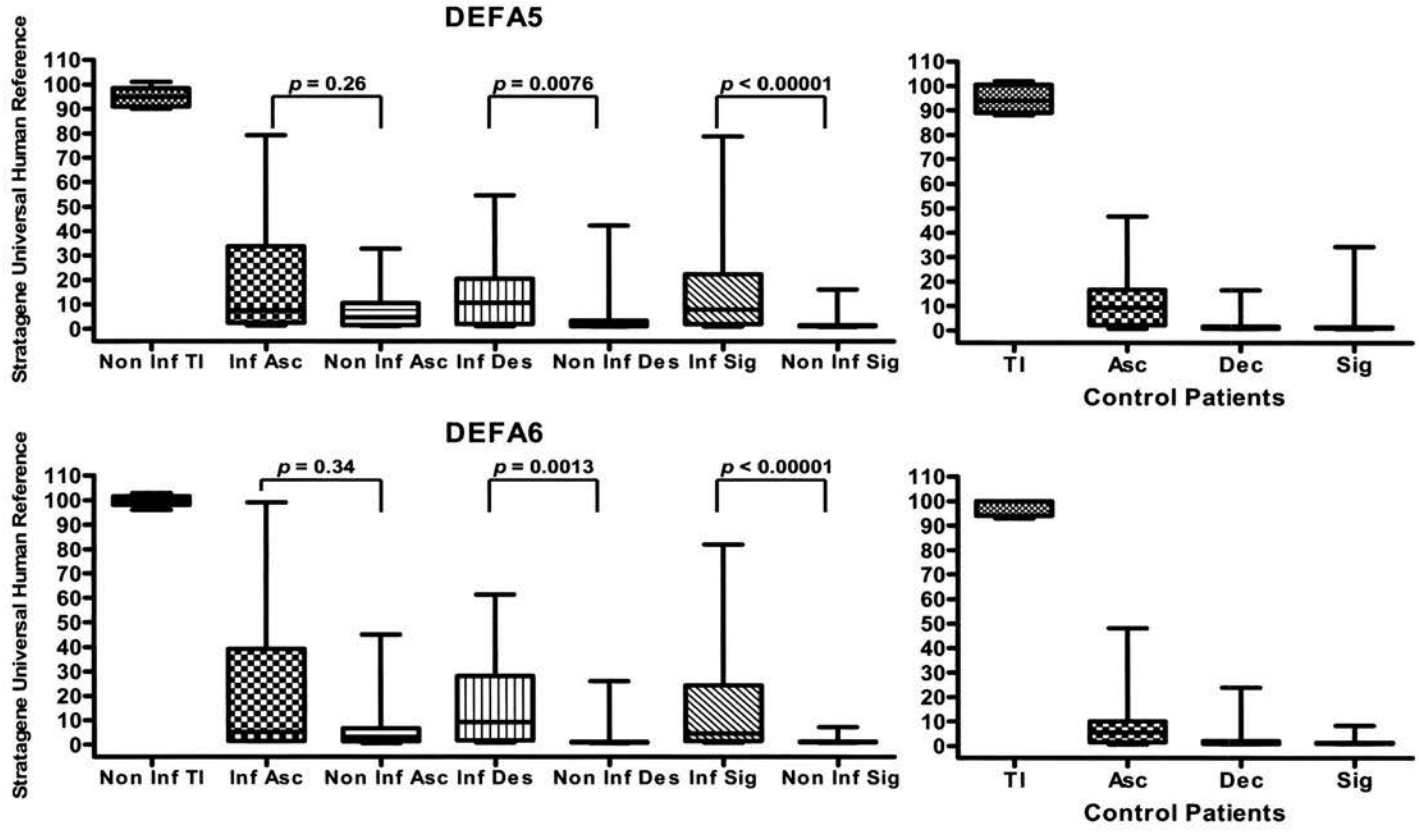
5.3.5 Analysis of Specific Gene Families- Alpha Defensins 5 and 6.

When DEFA5 and DEFA6 were analyzed expression in the normal controls and the non-inflamed ulcerative colitis biopsies was similar across the different anatomical locations with there being high expression in the terminal ileum, and expression decreasing as the biopsy location became more distal in the colon (Figure 5.2). In the acute and chronically inflamed UC biopsies there was marked upregulation of DEFA5 and DEFA6 expression in the descending and sigmoid colon (Table 5.4).

Table 5.6: Expression changes in the 40 most dysregulated sequences comparing inflamed ulcerative and non-inflamed ulcerative colitis sigmoid colon biopsies.

Gene Name(s)	Sequence Code	Fold Change	P-value
SAA1	A_24_P335092	16.50126	<10E-45
TNIP3	A_23_P386478	10.52272	1.03E-38
DEFA5	A_23_P112086	8.43598	<10E-45
MMP3	A_23_P161698	8.14743	2.33E-35
IL8	A_32_P87013	7.24295	8.42E-19
S100A9	A_23_P23048	7.11456	1.96E-32
regenerating islet-derived 3 gamma	A_32_P65628	6.98725	<10E-45
S100A8	A_23_P434809	6.83944	1.16E-19
KCND3	A_32_P140268	6.74598	6.27E-30
DEFA6	A_24_P363711	6.72245	4.16E-19
GW112	A_23_P2789	6.14232	1.95E-27
MMP7	A_23_P52761	5.5287	1.01E-23
LCN2	A_23_P169437	5.17756	1.30E-22
CXCL3	A_24_P183150	5.08356	2.26E-25
C14orf81	A_24_P608268	4.3061	2.78E-10
CHRDL2	A_23_P13548	4.08137	1.47E-16
CXCL1	A_23_P7144	4.05835	9.93E-13
REG4	A_24_P58673	3.99842	8.18E-21
DAF	A_23_P103951	3.81241	5.06E-37
CXCL10	A_24_P303091	3.74657	1.37E-14
TTTY15	A_24_P348861	-2.49501	7.06E-06
SLC26A2	A_23_P250951	-2.57291	1.25E-10
similar to MGC9515 protein	A_32_P215143	-2.62964	5.44E-09
A_32_P216970	A_32_P216970	-2.65464	6.22E-09
POSTN	A_24_P347411	-2.69014	6.46E-09
LOC63928	A_23_P349463	-2.71812	2.28E-12
PNLIPRP2	A_23_P24083	-2.7412	2.27E-27
ANKRD17	A_24_P220771	-2.76491	3.22E-11
MT1K	A_23_P66241	-2.77076	1.31E-13
CKB	A_23_P25674	-2.79887	3.67E-16
ATF3	A_24_P33895	-2.95597	0.00067
AQP8	A_23_P26522	-3.18702	3.51E-15
OSTalpha	A_24_P385732	-3.23873	6.36E-12
PCK1	A_23_P408249	-3.35965	3.86E-12
EIF1AY	A_24_P237511	-3.71062	2.30E-07
CLDN8	A_23_P427014	-3.712	9.70E-10
FLJ21934	A_24_P334378	-4.79656	0.00002
LOC285189	A_32_P197825	-5.02416	0.00355
LOC389023	A_32_P86578	-5.04036	<10E-45
SLC38A4	A_24_P321581	-5.73668	5.40E-14

Figure 5.2: Expression of defensins alpha 5 and 6 in ulcerative colitis patients and controls



Gene expression is shown as a box- whisker plot. Each endoscopic biopsy has been separated by patient status, inflammation status and anatomical location- TI- terminal ileum, Asc- ascending colon, Des- descending colon, Sig- sigmoid colon. Significantly higher DEFA5 and DEFA6 expression was observed in the inflamed UC descending and sigmoid colon biopsies when compared to the non-inflamed UC biopsies and the control biopsies.

5.3.6 ATP- binding cassette (ABC) transporter family and the Xenobiotic-transcription regulators

Expression patterns from probes representing 48 ABC transcriptional genes and their key mediators Pregnane X receptor (PXR), Farnesoid X-activated receptor, and Oxysterols receptor LXR-beta were analysed. When these genes were compared in all the ulcerative colitis and control biopsies, 7 genes were found to be significantly down regulated in the ulcerative colitis samples when compared to the control samples- ABCA1 ($p= 0.01$), ABCA8 ($p= 0.0064$), ABCB1 ($p= 0.00091$), ABCC6 ($p=0.0050$), ABCB7 ($p= 0.0068$), ABCF1 ($p=0.0005$) and ABCF2 ($p<0.00001$). Only one probe representing ABCB2 was significantly upregulated in ulcerative colitis ($p= 0.0048$).

ABCB1 expression was also significantly downregulated when inflamed ulcerative colitis sigmoid colon biopsies were compared to non- inflamed ulcerative colitis sigmoid biopsies. (FC-1.82, $p= 5.6x10^{-6}$) (Table 5.3). No difference in the expression of PXR between ulcerative colitis and controls was observed in any of the analysis.

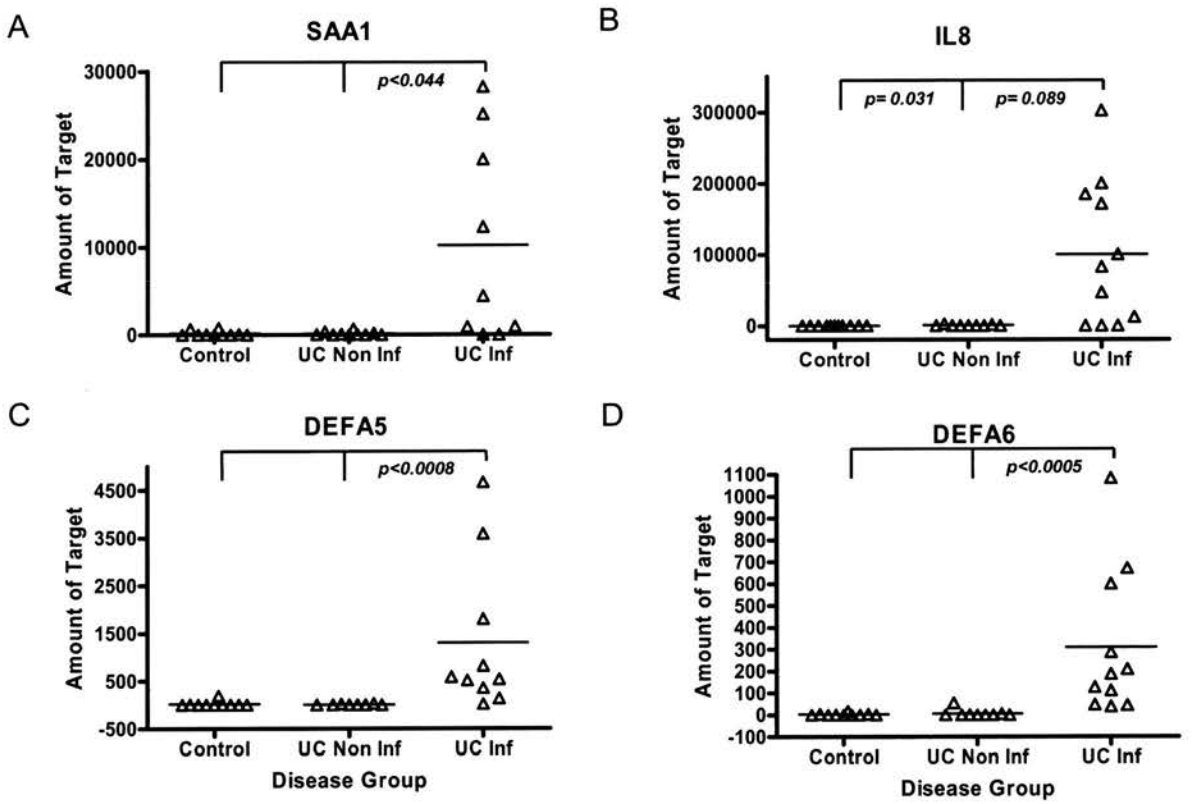
5.3.7 RTPCR Analysis

In 8 genes implicated by microarray expression results, confirmatory real time PCR analysis was undertaken in biopsies of patients form the original cohort following stratifying to represent a range of SAA1 and IL-8 expression. Increased SAA1 expression in the inflamed ulcerative colitis sigmoid colon biopsies compared to the normal control sigmoid colon

biopsies and the non-inflamed ulcerative colitis sigmoid colon biopsies (p=0.041 and p=0.044 respectively) was observed. Elevated IL-8 expression was also confirmed in the inflamed ulcerative colitis sigmoid biopsies when compared to the control sigmoid biopsies (p=0.031) (Figure 5.3).

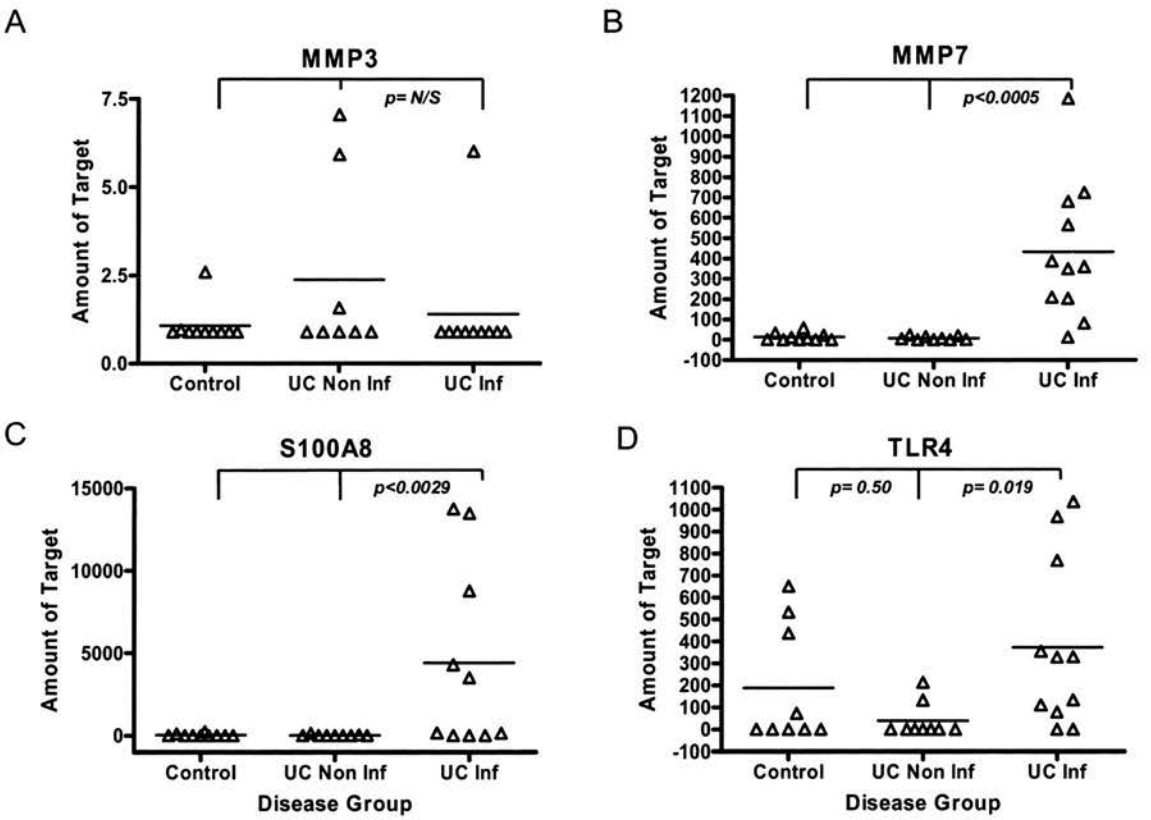
Increased expression of DEFA5 and DEFA6 in the inflamed ulcerative colitis sigmoid colon biopsies when compared to the non- inflamed ulcerative colitis sigmoid colon biopsies (p=0.0008 and p=0.0005 respectively) and the control sigmoid colon biopsies (p=0.0002 and p=0.0001 respectively) was observed (Figure 5.3). Increased expression in the inflamed ulcerative colitis sigmoid colon biopsies when compared to the non-inflamed ulcerative colitis sigmoid colon biopsies was also observed when MMP7, (p=0.0005), S100A8, (p=0.0029) and TLR4, (p=0.019) were examined (Figure 5.4).

Figure 5.3: Real time PCR expression of SAA1, IL-8 defensin alpha 5 and defensin alpha 6 in control and ulcerative colitis inflamed and non-inflamed sigmoid colon biopsies.



Real time PCR expression data comparing expression in 10 healthy control sigmoid biopsies with normal histology, 9 quiescent UC sigmoid biopsies and 11 UC sigmoid biopsies with an acute or chronic inflammatory cell infiltrate. Expression of SAA1, IL-8, DEFA5 and DEFA6 were compared between the control and the inflamed and non-inflamed UC biopsies.

Figure 5.4: Real time PCR expression of MMP3, MMP7, S100A8 andTLR4 in control and ulcerative colitis inflamed and non- inflamed sigmoid colon biopsies.



Expression of MMP3, MMP7, S100A8 and TLR4 were compared between the different patient groups.

5.3.8 *In-Situ* Hybridization and Immunohistochemistry

In- Situ hybridization of the terminal ileal biopsies for DEFA5 showed strong hybridization in the basal crypts consistent with Paneth cell location (Figure 5.5). In the ulcerative colitis biopsies taken from the sigmoid colon strong, multifocal hybridization in the basal crypt region of these biopsies was observed and this would be consistent with Paneth cell metaplasia. This was not observed in the non-inflamed control biopsies.

Immunohistochemistry for DEFA6 confirmed that in the sigmoid colon ulcerative colitis biopsies, staining was observed in the basal crypt region of these biopsies consistent with Paneth cell metaplasia Again, this was not observed in the control biopsies (Figure 5.6).

5.3.9 Expression of Genes within the IBD2 Locus

Using the markers defining the IBD2 locus we identified 526 *Agilent* probes representing genes or expressed sequence tags within this locus on chromosome 12. 12 probes had a greater or less than 1.5 fold change in expression with $p < 0.01$ when expression of acute and chronically inflamed ulcerative colitis sigmoid colon biopsies were compared to non-inflamed ulcerative colitis sigmoid colon biopsies (Table 5.7).

Figure 5.5: Defensin alpha 5 In- Situ Hybridization in the terminal ileum and colon of patients with ulcerative colitis and controls.

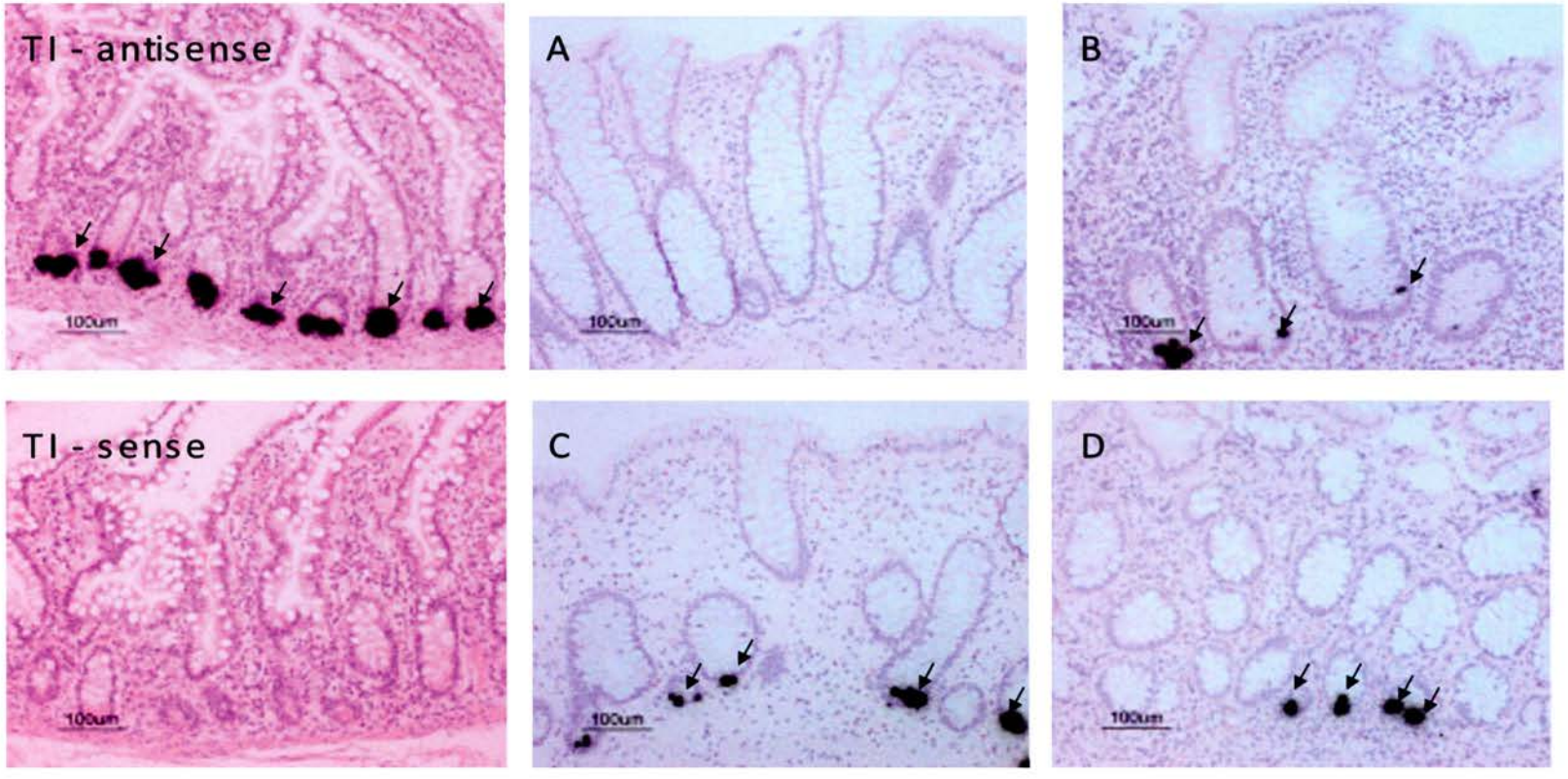


Figure 5.5 Terminal ileum (TI) upper panel: antisense probe shows strong hybridization in the basal crypts consistent with Paneth cell location. Arrows indicate positive staining in crypt cells. Lower panel: no significant hybridization with sense control probe. (A) Sigmoid colon biopsy of a non-inflamed control patient. (B,C& D) Strong, multifocal hybridization in the basal crypt region of sigmoid colon biopsies from 3 ulcerative colitis patients consistent with Paneth cell metaplasia.

Figure 5.6: Defensin alpha 6 Immunohistochemistry in the terminal ileum and colon of patients with ulcerative colitis and controls.

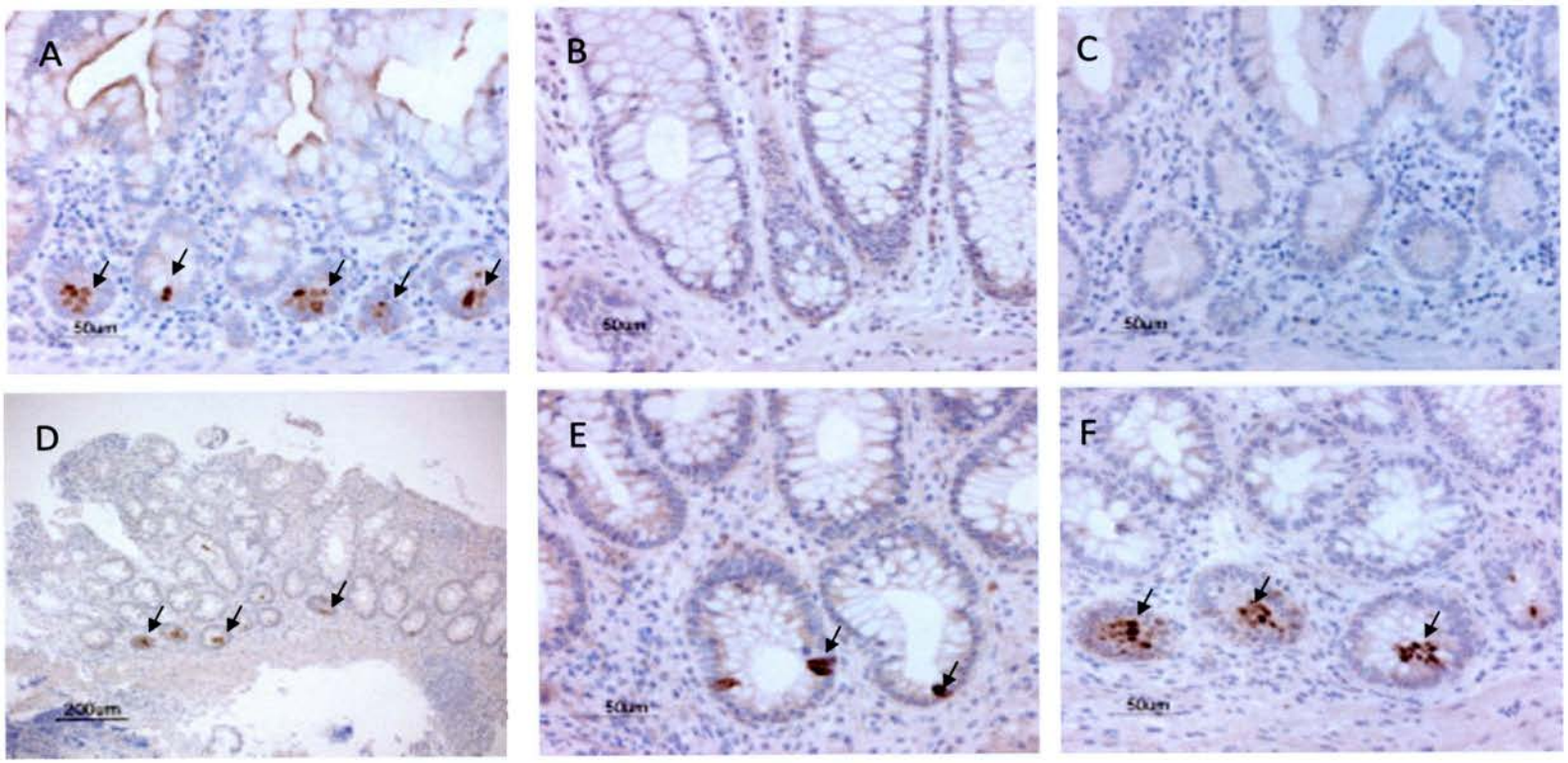


Figure 5.6 (A) Terminal ileal immunohistochemistry shows positive staining marked with arrows in the basal crypts consistent with Paneth cell location. (B & C) No significant staining was observed in two non-inflamed control patients. (D, E & F) Strong, multifocal staining in the basal crypt region of sigmoid colon biopsies from 3 ulcerative colitis patients consistent with Paneth cell metaplasia.

Table 5.7: Gene expression in probes from the IBD2 locus, comparing inflamed ulcerative colitis sigmoid colon biopsies to non-inflamed ulcerative colitis sigmoid colon biopsies.

Agilent Probe	Gene Symbol	ulcerative colitis sigmoid inflamed (35 biopsies) v non-inflamed (25) (FC)	p value
A_23_P98876	SLC39A5	-1.52855	$1.36x10^{-7}$
A_24_P647146	HDAC7A	1.53876	$3.15x10^{-16}$
A_24_P941773	DKFZP586A0522	-2.27306	$2.31x10^{-11}$
A_24_P945113	ACVRL1	1.57956	$6.33x10^{-9}$
A_23_P128230	NR4A1	1.81844	0.00005
A_23_P331098	K5B	2.3845	0.0004
A_24_P246636	A_24_P246636	-1.69754	$7.06x10^{-8}$
A_23_P2233	SILV	1.51557	$9.07x10^{-9}$
A_23_P105251	GLI	-1.54447	$7.73x10^{-9}$
A_32_P3783	HMGA2	-1.94107	$2.37x10^{-9}$
A_23_P162300	IRAK3	1.7382	$3.11x10^{-15}$
A_32_P83256	IRAKM	1.70847	$4.56x10^{-10}$

Analysis of the 526 expression probes located within the IBD2 locus identified 12 probes that were significantly differentially regulated when the inflamed ulcerative colitis sigmoid colon biopsies were compared to the non- inflamed ulcerative colitis sigmoid colon biopsies. Gene annotation- SLC39A5- solute carrier family 39A5, HDAC7A- histone deacetylase 7A, DKFZP586A0522- Methyltransferase-like protein 7A precursor, ACVRL1- activin A receptor type II-like 1, NR4A1- nuclear receptor subfamily 4, group A, member 1, KB5- keratin 5B, A_24_P246636- unknown, SILV- silver homolog (mouse), HMGA2- high mobility group AT-hook 2, IRAK3- interleukin-1 receptor-associated kinase 3, IRAKM- interleukin-1 receptor-associated kinase M.

5.4 DISCUSSION

In the present study we have taken particular care to address concerns currently being expressed about micoarray studies in disease. We initially documented the regional variation of gene expression in the healthy colon before undertaking comprehensive studies in ulcerative colitis. By documenting the anatomical location for each biopsy, using appropriate inflammatory controls, as well as the avoidance of pooling of samples, we have been able to remove a considerable amount of background variability that has hampered previous studies. (157;169;174;179) A further strength of our study has been the fact that real time PCR analysis consistently confirmed the significant changes in expression, substantially increasing the confidence associated with the interpretation of the data.

We have made a number of novel observations. In the healthy adult colon this is the first microarray study to show a gradient of expression of a number of genes. Genes involved in developmental pathways- the HOX family and the hedgehog signaling pathway appeared to be the most differentially regulated along the anatomical length of the healthy colon. HOXA13 has been shown to play a crucial role in the development of tail gut and mutations in the gene result in urogenital abnormalities,(240) and interestingly it has been shown that HOXB13 expression is down regulated in colorectal tumours from the distal left colon.(241)

GLI1 is one of the major effector molecules of the hedgehog signaling pathway and the GLI1 gene lies within the IBD2 locus, strongly implicated in ulcerative colitis.(79;242;243) Data from the present study would also suggest that GLI1 is downregulated in inflamed ulcerative

colitis biopsies compared to non-inflamed biopsies from patients with ulcerative colitis. These data add further weight to data from Cambridge and our own unit showing an association between mutations in the GLI1 gene and ulcerative colitis.(244)

With regards to the observed gradient of expression in the healthy adult colon, our data contrast with data from Costello and Wu and colleagues where no significant differences in expression patterns were observed when comparing biopsies from caecum, transverse colon, descending colon and sigmoid colon.(196;245) The observed differences in these data sets may be explained by the fact that we have considered only non- inflamed healthy controls in these studies and have not pooled healthy and diseased data. When anatomical variation in the colon was compared between ulcerative colitis and control patients there was only a 34.5% homology between the differentially expressed gene ontology groups. This difference may be explained by environmental factors in the ulcerative colitis colon such as microbial dysbiosis.

In ulcerative colitis we demonstrated dysregulation of genes involved in innate immunity, notably the alpha defensins 5 and 6, together with other pathways recurrently implicated in inflammatory bowel disease. In our data set high levels of alpha defensin 5 and 6 expression were observed in the terminal ileal biopsies of non-inflamed controls and patients with ulcerative colitis. Levels of expression in these patients fell as the location that the biopsies were retrieved from became more distal in the colon- ascending colon, descending colon and sigmoid colon. However, in the inflamed ulcerative colitis biopsies increased expression of both alpha defensins 5 and 6 was observed in the descending and sigmoid colon. Lawrance and colleagues also noted that the defensins alpha 5 and 6 were upregulated in patients with

UC compared to controls,(167) although RNA was extracted from surgical resections and no details about the anatomical location of these specimens were given.

Recent data published by Varnat and colleagues have suggested that PPARβ negatively regulates Paneth cell differentiation by downregulating the expression of Indian hedgehog, another of the major effector molecules in the hedgehog signaling pathway.(246) Immunohistochemistry and *in-situ* hybridization have shown that this is largely mediated by Paneth cell metaplasia. Given our present data now implicating germline GLI1 variation in disease susceptibility, and the regional variation in health, we speculate that in patients with ulcerative colitis, further as yet undetermined defects in the Hedgehog signaling pathway may result in unregulated Paneth cell differentiation, Paneth cell metaplasia, increased alpha defensin 5 and 6 expression, and mucosal inflammation.

It is of interest that many but not all of our results are broadly in line with two of the landmark microarray papers in inflammatory bowel disease. Consistent with data from Lawrance and colleagues (167) we have shown upregulation of S100A8 & A9 and the alpha defensins 5 & 6 in ulcerative colitis. Dieckgraefe and colleagues (166) observed upregulation of a number of the MMP genes, again seen in our dataset. Another consistent finding was the upregulated expression of members of the REG family in the colon of patients with ulcerative colitis, probably as a result of Paneth cell metaplasia.(247)

The downregulation of ABCB1 in our dataset is of significant interest, and consistent with earlier microarray data from Langmann, Dieckgraefe, Lawrance, Wu and colleagues. (166-

168;245) It is pertinent that when the entire class of proteins sharing homology with ABCB1 were analyzed, a further 6 out of 48 ABC transporters were significantly dysregulated in UC, including ABCA1, ABCA8, ABCC6, ABCB7, ABCF1 and ABCF2. An important role of the ABC transporters in the aetiopathogenesis of UC seems likely, supported by these consistent microarray data, association with germline MDR1 variability, and animal data to date.

However, and in contrast to data produced by Langmann we did not observe any changes in expression of the transcriptional regulator Pregnane-X receptor. These negative data are consistent with genetic studies carried out in the inflammatory bowel disease population in Edinburgh- using a haplotype tagging approach, there was an association between the ABCB1 gene and ulcerative colitis,(248) but no association between the Pregnane- X receptor and ulcerative colitis.(249) Aspects of study design and patient recruitment may explain the differences observed between our data and those of Langmann and colleagues.

When novel genes were considered, of particular note was the differential expression of the poorly characterized gene TNFAIP3-interacting protein 3 (TNIP3) which is a nuclear or cytoplasmic protein with three coiled domains that was first pulled down from a Yeast-2 hybrid scan of TNFAIP3 (A20).(250) The protein has been found to be expressed in macrophages and is upregulated by infection with *Listeria.* Further to this TNFAIP3 was recently identified as the closest gene to rs7753394 a 'second tier' hit from the UK genome wide scan in Crohn's disease, focusing further interest on this NFκB dependant signaling pathway.(57)

In conclusion these data provide a rigorously characterized expression profile of the whole genome in the terminal ileum and colon of patients with ulcerative colitis and controls. The studies provide new insights into regional variation of gene expression in the healthy colon, and also considerably extend previous studies in ulcerative colitis. These data also identify a number of key regulators of intestinal inflammation worthy of further study. As further data from genome-wide scanning emerges in this and other complex diseases, access to these data and the ability to study expression, function and germline variation in parallel will become all the more necessary.

CHAPTER 6

THE CONTRIBUTION OF OCTN1/2 VARIANTS WITHIN THE IBD5 LOCUS TO DISEASE SUSCEPTIBILITY AND SEVERITY IN CROHN'S DISEASE

SUMMARY

Background & aims: Peltekova and colleagues have suggested that polymorphisms in the organic cation transport genes, OCTN1 (SLC22A4) and OCTN2 (SLC22A5) represent the disease-causing mutations within the IBD5 locus (chromosome 5q31). We investigated associations with disease susceptibility, phenotype, and evidence for epistasis with CARD15 in 679 patients with Crohn's disease (CD) or ulcerative colitis (UC). Expression of probes representing 11 genes within the IBD5 locus was also investigated in endoscopic colonic and terminal ileal biopsies.

Methods: 374 CD, 305 UC and 294 controls (HC) were studied. Genotyping for SNPs IGR2096, IGR2198, IGR2230, OCTN1 variant (SLC22A4 1672C→T) and OCTN2 variant (SLC22A5 -207G→C) was performed using the Taqman system. Gene expression was compared in 53 CD, 67 UC and 31 control subjects.

Results: The IBD5, OCTN1 and OCTN2 polymorphisms were in strong linkage disequilibrium (D' >0.959). IGR2198 variant allele frequency (49.1% v 40.8%,P=0.0046) and homozygosity (21% v 14.8%,P=0.044) were associated with CD v HC. Variant allelic frequency of OCTN1 (53.6% v 43%,P=0.0008), OCTN2 (56.1% v 48.4%,P=0.0092) polymorphisms and homozygosity for the OCTN1/2 TC haplotype (28.4% v 16%,P=0.0042) were associated with CD v HC. IGR2198 homozygosity and TC homozygosity were associated with stricturing/penetrating disease at follow-up (P=0.011, P=0.011 respectively) and disease progression (P=0.038, P=0.049) on univariate analysis, and with need for surgery on multivariate analysis (P=0.006, P=0.004). In the absence of the IBD5 risk haplotype, no association of OCTN1/2 variants with CD was detected. No associations were seen with UC. OCTN2 expression was marginally downregulated in CD and UC biopsies compared to controls.

Conclusions: The IBD5 locus influences susceptibility, progression and need for surgery in CD. However, the contribution of OCTN1/2 variants is not independent of the IBD5 haplotype: a causative role for these genes remains plausible but is not yet proven.

6.1 INTRODUCTION

The 5q31-33 area was originally identified in 1999 as conferring susceptibility for Crohn's disease.(75) This finding was replicated using linkage disequilibrium mapping in the Canadian Crohn's disease population and further delineated to the IBD5 loci- 5q31.(76) Fine mapping of this area has identified a single, highly conserved 250-kb haplotype of 11 SNPs spanning a cytokine gene cluster that is associated with Crohn's disease.(77) Further high resolution analysis of the 5q31 region using 103 SNPs revealed 11 discrete haplotype blocks that measure tens of kilobases in length, have limited diversity, and are punctuated by sites of recombination.(89)

Peltekova and colleagues have suggested that mutations of two genes within the IBD5 region, the organic cation transporters OCTN1 (SLC22A4) and OCTN2 (SLC22A5), may be independently associated with Crohn's disease.(95) The construction of a two allele risk haplotype OCTN1/2 TC (SLC22A4 exon 9 1672C→T and SLC22A5 promoter -207G→C) was reported to be associated with Crohn's disease in patients who lacked the extended IBD5 risk haplotype.

Peltekova and colleagues also suggested that the OCTN1 variant (1672C→T) alters its function in fibroblasts in vitro, with variant forms having less affinity for carnitine and a greater affinity for tetraethyl ammonium and some xenobiotics, and that the OCTN2 variant (-207G→C) disrupted a heat-shock transcription factor binding site in fibroblasts in vitro.(95)

In the present study we have assessed the contribution of the OCTN1 and OCTN2 polymorphisms implicated by Peltekova and colleagues in determining genetic susceptibility in Crohn's disease and ulcerative colitis, specifically addressing whether these OCTN1/2 variants have an association with susceptibility to inflammatory bowel disease that is independent of other markers within the extended the IBD5 linkage interval. We have also investigated whether these polymorphisms are associated with specific disease phenotype, assessed gene-gene interactions with established NOD2/CARD15 mutations and examined expression of 11 probes representing genes with the IBD5 locus in human colonic and terminal ileal biopsies.

6.2 METHODS

6.2.2 Patients and Controls

Six hundred and seventy nine patients with well characterized inflammatory bowel disease (IBD) and 294 control patients were recruited (Table 6.1). The IBD group comprised 374 patients with Crohn's disease (CD) and 305 patients with ulcerative colitis (UC). Classification and data collection of these patients are described in the methods section (Methods 3.4.1-3). Disease extent was recorded at diagnosis. The controls comprised of 294 healthy subjects and involved blood donors (163) recruited from the south east of Scotland and healthy control subjects (131). There were 143 males and 151 females with a median age of 39 years (IQR 27-52).

6.2.2 Genotyping

To examine the relative contribution of the OCTN1/2 variants relative to the extended IBD5 haplotype, 3 SNPs were genotyped- IGR2096 (which lies within haplotype block 4, as defined by Daly et al (89)) (Figure 6.1), IGR2198 (within haplotype block 5) and IGR2230 (within haplotype block 7). The rs1050152 polymorphism of the OCTN1 gene (SLC22A4 exon 9 1672C→T, IGR3002) and the rs26313667 (SLC22A5 promoter, -207G→C, IGR2222) polymorphism of the OCTN2 gene were genotyped (Table 6.2). CD patients and controls were genotyped for polymorphisms of the NOD2/CARD15 gene (R702W, G908R, and 1007fsinsC) using previously described methods.(85)

Table 6.1: Demographics and clinical features of the Crohn's disease, Ulcerative colitis and control groups.

	Crohn's Disease	Ulcerative colitis	Controls
Total Number	**374**	**305**	**294**
Sex (male / female)	181/193	171/134	143/151
Median Age at diagnosis (yrs) (IQ range)	27.8 (20.9-40.4)	34 (25-50)	39(27-52)
Median duration of follow up (yrs) (IQ)	11.8 (6.5-20.2)	7.5 (3.35-13.4)	
Caucasian (%)	98.7%	98.3%	
Age at diagnosis A1(<40 years)/ A2(>40)	72% / 28%		
Location at Diagnosis			
Ileal disease (L1)	125 (34%)	Proctitis 105 (34.5%)	
Colonic Disease (L2)	139 (38%)	Left sided colitis 116 (37%)	
Ileal and colonic (L3)	58 (16%)	Extensive colitis 84 (27.5%)	
Upper GI disease (L4)	30 (8%)		
Oral Crohn's disease	6 (2%)		
Peri-anal disease	75 (21%)		
Disease Behaviour at Diagnosis			
Inflammatory (Vienna B1)	258 (74%)		
Stricturing (Vienna B2)	32 (9%)		
Penetrating (Vienna B3)	61 (17%)		
Disease Behaviour at Follow Up			
Inflammatory (B1)	128 (34%)		
Stricturing (B2)	70 (19%)		
Penetrating (B3)	168 (48%)		
Disease Progression			
No progression from inflammatory (% of inflammatory at diagnosis)	126 (49%)		
Inflammatory to stricturing or penetrating.	132 (51%)		
Surgery for luminal complications of CD[1]	237 (63%)		

Disease behaviour at follow up was defined using the Vienna classification when the patient was last clinically evaluated. Full phenotypic data were available on 94% of the CD patients at diagnosis and 98% at follow up. In assessing disease progression, patients were grouped as those who remained as inflammatory (B1), non progressive disease and those whose disease had progressed to stricturing (B2) or penetrating (B3) disease.

[1] Examination under anaesthesia or drainage procedures for peri-anal sepsis were excluded.

Table 6.2: SNP nomenclature and primers that were used for genotypic analysis.

IGR Number	Registered SNP number	Gene represented	Taqman Primers
IGR2096	rs12521868		Forward: tctgagacaggagccactagag Reverse: cacagcatccagagtgatcct
IGR2198	rs11739135		Forward: gggttgcatgagcattaagtttcaa Reverse: ccacatcaaggataagactgctaaatact
IGR2222	rs26313667	OCTN2 (SLC22A5)	Forward: gcggctggccttacatagg Reverse: ccgctctgcctgcca
IGR2230	rs17622208		Forward: gcaggcagaacagccatact Reverse: ggccacagaactttcattaaagtagga
IGR3002	rs1050152	OCTN1 (SLC22A4)	Forward: tttacaggtgcttacaacagaatg Reverse: tagtctgactgtcctgattggaatc

6.2.3 Expression of Genes Within the IBD5 Locus

Gene expression was analyzed in the IBD5 locus using data generated from the microarray gene expression study. Detailed methods are provided in the methods section 3.2.1-9, 3.3.4. Eleven probes were identified that represented genes within the IBD5 locus. Statistical significance of the microarray data was determined by Student's unpaired *t* test. $p < 0.01$ and a fold change of greater or less than 1.5 were considered statistically significant.

6.2.4 Data Analysis

Each SNP was analyzed for association with inflammatory bowel disease overall, Crohn's disease, ulcerative colitis and disease phenotype. Each of the variants studied was shown to be in Hardy-Weinberg equilibrium in patients, and in controls. Evidence for epistasis between the OCTN1/2 allelic variants, the OCTN1/2 TC haplotype and NOD2/CARD15 variants was investigated by comparing allelic frequencies of the individual OCTN1/2 variants, together

with homozygosity for the OCTN1/2 TC haplotype between the sub-groups of patients with and without NOD2/CARD15 variants by Chi squared analysis.

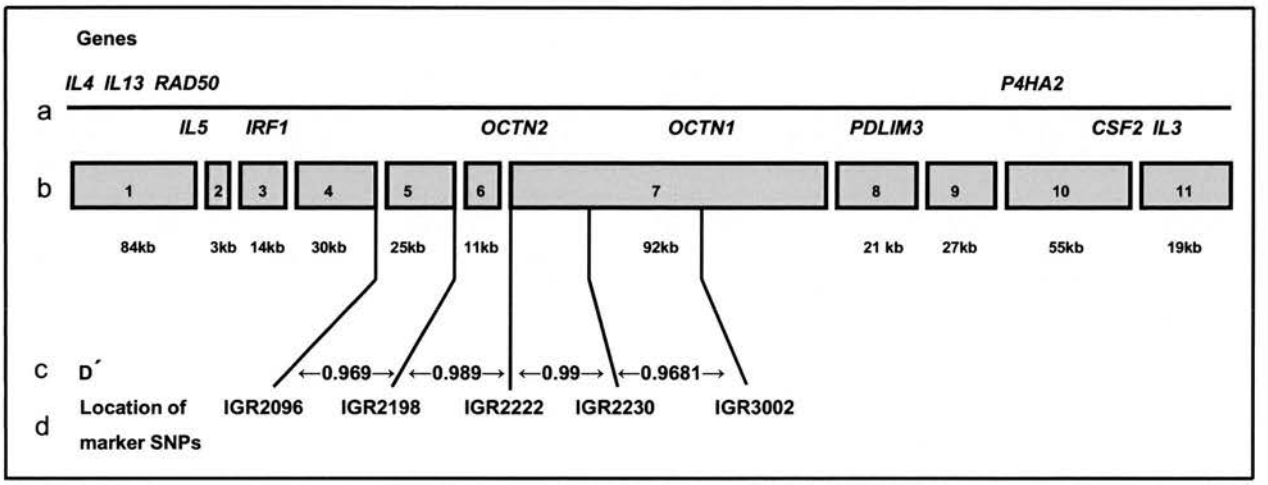
To identify significant independent variables associated with genotype, multiple logistic regression analysis was carried out using Minitab v10. The population attributable risk percentage (PAR%) was estimated by the method of Schlesselman.(215) To calculate this, the prevalence of Crohn's disease was estimated at 100/100,000 and the frequency of all alleles in the control population was assumed to reflect that of the general population.

6.3 RESULTS

6.3.1 Linkage Disequilibrium across the IBD5 Locus

In the Crohn's disease population, strong linkage disequilibrium was observed between the SNPs within the extended IBD5 haplotype as defined by Daly and colleagues (89)-IGR2096 (block 4), IGR2198 (block 5), OCTN2 (SLC22A5 promoter, -207G→C) (block 7), IGR2230 (block 7) and OCTN1 (SLC22A4 exon 9 1672C→T)(block 7)(Figure 6.1). Pairwise D′ values (a measure between 0 and 1 where 0 is equal to no correlation and 1 shows complete correlation) of greater than 0.959 between each SNP confirmed the difficulty of showing an independent effect of OCTN1/2 variants and IBD5. In the detailed analyses of the IBD5 contribution to disease phenotype presented in the manuscript, data are presented for the IGR2198 marker, representative of the three-locus haplotype (IGR2096, IGR2198, IGR2230).

Figure 6.1. Haplotype Structure and Linkage Disequilibrium across the IBD5 region.



The IBD5 locus (5q31) with the high resolution haplotype structure as reported by Daly and colleagues (27): **a:** Candidate genes above the relevant haplotype blocks. Genes above the line are transcribed from left to right and those below the line are transcribed from right to left. **b:** Eleven blocks numbered 1 to 11 (between 3kb and 92kb) each of limited genetic diversity are punctuated by sites of recombination. **c:** D´ scores are shown to demonstrate the tight linkage disequilibrium between the SNPs that were analysed. The lowest D´ (D´ =0.959) was observed between IGR2096 and the OCTN2 variant (-207G→C). **d:** Location of the SNPs that were analysed, IGR2222 representing the OCTN2 variant (-207G→C) and IGR3002 representing the OCTN1 variant (1672C→T). IGR2078 which was used by Peltekova and colleagues to represent the extended IBD5 haplotype is located in block 4.

6.3 2 Analysis of SNPs representing the extended IBD5 haplotype (IGR2096, IGR2198 and IGR2230).

Variant allelic frequencies differed significantly between Crohn's disease (CD) and healthy controls (HC) for each of the three IBD5 single nucleotide polymorphisms studied, IGR2096 (49.4% CD v 41.9% HC, p= 0.030), IGR2198 (49.1% CD v 40.8% HC, p= 0.0046) and IGR2230 (54.9% CD v 47.4% HC, p= 0.011) (Table 6.3). Carriage rates for the IBD5 variants were significantly higher in the CD patients when compared to the control population IGR2096 (79.3% CD v 68.7% HC, p= 0.0034), IGR2198 (77% CD v 67% HC, p= 0.011) and IGR2230 (80% CD v 73.7% HC, p= 0.047). Individuals who were homozygous for the IBD5 risk alleles at the three SNPs examined (IGR2096, IGR2198 and IGR2230) were more common in the CD group than the control group (Table 6.3). The PAR for IBD5 homozygosity was estimated as 15% if IGR2096 data were considered in calculation, 14.0% for IGR2198, and 14.3% for IGR2230.

No associations were observed between allelic frequency of IBD5 variants and inflammatory bowel disease overall (IBD) (IGR2198, 45% IBD v 40.8% HC, p= 0.077) or ulcerative colitis (UC) (IGR2198, 43.7% UC v 40.8% HC, p= 0.325) (Table 6.4).

Table 6.3: IBD5 variant allele frequency, carriage rates and homozygote frequency in Crohn's disease patients and controls.

SNP / Haplotype Examined	Control Frequency	CD disease frequency	P value	RR (CI)
Variant Allelic Frequency				
IGR2096	211 (41.9%)	347 (49.4%)	0.015	1.42 (1.1-1.8)
IGR2198	209 (40.8%)	328 (49.1%)	0.0046	1.40 (1.1-1.6)
OCTN2 (-207 G→C)	242 (48.4%)	386 (56.1%)	0.0092	1.43 (1.2-1.8)
IGR2230	232 (47.4%)	355 (54.9%)	0.011	1.35 (1.1-1.7)
OCTN1 (1672C→T)	216 (42.9%)	356 (53.6%)	0.0008	1.48 (1.2-1.9)
Variant Carriage Rates				
IGR2096	173 (68.7%)	264 (79%)	0.0034	1.75 (1.2-2.6)
IGR2198	171 (67%)	258 (77%)	0.011	1.61 (1.2-2.3)
OCTN2 (-207 G→C)	187 (74%)	280 (81%)	0.034	1.52 (1.0-2.2)
IGR2230	185 (73.7%)	261 (80%)	0.047	1.5 (1.1-2.23)
OCTN1 (1672C→T)	178 (68.7%)	267 (80%)	0.0024	1.77 (1.2-2.6)
TC Haplotype	170 (69%)	264 (80%)	0.0016	1.8 (1.3-2.7)
IBD5 Haplotype[1]	166 (63.6%)	246 (73%)	0.016	1.5 (1.1-2.2)
Variant Homozygote Frequency				
IGR2096	38 (15.1%)	73 (21.4%)	0.036	1.58 (1.0-2.4)
IGR2198	38 (14.8%)	70 (21%)	0.044	1.56 (1.0-2.4)
OCTN2 (-207 G→C)	55 (22%)	106 (31%)	0.0155	1.59 (1.1-2.3)
IGR2230	47 (18.6%)	94 (29%)	0.0038	1.79 (1.2-2.6)
OCTN1 (1672C→T)	44 (17%)	89 (25.6%)	0.011	1.69 (1.1-2.5)
TC Haplotype	40 (16%)	86 (28.4%)	0.0042	1.83 (1.2-2.8)
IBD5 Haplotype[1]	38 (14.6%)	68 (20.5%)	0.071	1.48 (1.0-2.3)

Three single nucleotide polymorphisms (SNPs) were used to define the extended IBD5 locus (IGR2096, IGR2198 and IGR2230) (Figure 1). Each was independently associated with susceptibility to Crohn's disease when allelic frequencies, carriage rates and homozygosity were analysed. OCTN2 variant (-207G→C) and OCTN1 variant (1672C→T) were also independently associated with susceptibility to Crohn's disease when allelic frequencies, carriage rates and homozygosity were analysed. The two allele risk haplotype OCTN1/2 TC haplotype was associated with susceptibility to Crohn's disease when carriage rates, and homozygosity rates were analysed.

[1] A two allele risk haplotype using the IBD5 marker SNPs IGR2198 and IGR2230 is also illustrated for comparison to the OCTN1/2 TC haplotype.

Table 6.4: IBD5 variant allele frequency, carriage rates and homozygote frequency in ulcerative colitis.

SNP/ Haplotype Examined	Frequency in healthy controls n=294	Frequency in UC patients n=305	P value (RR)
Allelic Frequency			
IGR2198	209 (40.8%)	266 (44.2%)	0.26 (1.1)
OCTN2 (-207 G→C)	242 (48.4%)	295 (51%)	0.38 (1.1)
OCTN1 (1672C→T)	216 (42.9%)	270 (44.7%)	0.53 (1.1)
Carriage Rates			
IGR2198	171 (67%)	204 (68%)	0.86 (1.0)
OCTN2 (-207 G→C)	187 (74%)	217 (72%)	0.56 (0.9)
OCTN1 (1672C→T)	178 (68.7%)	205 (68%)	0.98 (0.97)
TC Haplotype	170 (69%)	196 (65%)	0.4 (0.86)
Homozygote Frequency			
IGR2198	38 (14.8%)	62 (20.6%)	0.11 (1.43)
OCTN2 (-207 G→C)	55 (22%)	78 (26%)	0.31 (1.23)
OCTN1 (1672C→T)	44 (17%)	65 (22%)	0.26 (1.28)
TC Haplotype	40 (16%)	63 (21%)	0.1 (1.38)

No significant associations were observed in ulcerative colitis patients between disease susceptibility and any variant studied.

6.3.3 Analysis of OCTN1/2 variants and OCTN1/2 TC haplotype

Allelic frequencies differed between Crohn's disease patients and controls for the OCTN1 variant (SLC22A4 exon 9 1672C→T) (CD 53.6% v HC 42.9%, p= 0.0008) and the OCTN2 variant (SLC22A5 promoter, -207G→C) (CD 56.1% v HC 48.4% p= 0.0092) (Table 6.3). Carriage of the OCTN1/2 TC risk haplotype was present more frequently in patients with CD than controls (80% CD v 68.5% HC, p= 0.0016, RR=1.8). It was clear that this difference related to homozygosity as OCTN1/2 TC homozygotes were more common in the Crohn's disease group (28.4% CD v 16.1% HC, p= 0.0042, RR 1.83). No significant difference was observed between OCTN1/2 TC heterozygote rates (p= 0.3).

An association was observed between the OCTN1 variant and inflammatory bowel disease when allelic frequencies were analyzed (IBD 48.9% v HC 42.9%, p= 0.019). This finding was not replicated in analysis of the OCTN2 variant (IBD 51.7% v HC 48.4%, p= 0.162) and no associations were observed between variant allelic frequencies of OCTN1 (UC 44.5% v HC 42.8%, p= 0.58) or OCTN2 (UC 48.1% v HC 47.8%, p= 0.92) and ulcerative colitis (Table 6.4). No association was observed between the OCTN1/2 TC haplotype and IBD overall, or with UC. There was no evidence of epistasis between the OCTN1/2 variants, the OCTN1/2 TC haplotype and CARD15 variants (Table 6.5).

Table 6.5: Frequency of OCTN1/2 variant alleles and the OCTN1/2 TC haplotype in Crohn's disease patients stratified by carriage of NOD2/CARD15 variants.

SNP / Haplotype Examined	NOD2/CARD15 positive patients n=84	NOD2/CARD15 negative patients n=188	P value
OCTN2 (-207 G→C) Allelic Frequency (%)	55.2	56.3	0.78
OCTN1 (1672C→T) Allelic Frequency (%)	51.9	53.4	0.69
OCTN1/2 TC Haplotype Carriage (%)	72.3	71.1	0.79
OCTN1/2 TC Haplotype Homozygosity (%)	22.0	24.5	0.574

Analysis of evidence for epistasis between OCTN2 variant (-207G→C) and OCTN1 variant (1672C→T) and the TC haplotype of OCTN1/2 was investigated by stratifying OCTN1/2 variants by carriage of one or more of the three common CARD15 variants- R702W, G908R, and 1007fsinsC. No evidence of epistasis was observed.

6.3.4 The OCTN1/2 association with Crohn's Disease is not independent of the association with the extended IBD5 haplotype.

Previous data have suggested carriage of the OCTN1/2 TC haplotype to be an independent risk factor for Crohn's disease. However, in our data-set, individuals who lacked the IBD5 risk haplotype (homozygous with respect to the non- Crohn's disease associated alleles of IGR2096, IGR2198 and IGR2230) the OCTN1/2 TC haplotype was not associated with Crohn's disease (Table 6.6). In the absence of variants in the marker SNP IGR2096, which is located in block 4, 21.5% of the controls carried the OCTN1/2 TC haplotype compared to 30.4% in the Crohn's disease group (p= 0.22). When the OCTN1/2 TC haplotype was analysed in the absence of IGR2198 variants (block 5) 27.1% of the controls possessed the OCTN1/2 TC haplotype versus 17.1% of the Crohn's disease patients (p= 0.13). In the absence of variants of the SNP IGR2230 which is in the same block as OCTN1/2 (block 7) 3.2% of the controls possessed the OCTN1/2 TC haplotype versus 0% of the Crohn's disease patients (p= 0.25).

Table 6.6: Frequency of the OCTN1/2 TC haplotype in individuals not possessing disease susceptibility risk alleles at the markers IGR2096, IGR2198, and IGR2230.

	No of subjects not possessing IGR2096 variants	OCTN1/2 TC Haplotype frequency	No of subjects not possessing IGR2198 variants	OCTN1/2 TC Haplotype frequency	No of subjects not possessing IGR2230 variants	OCTN1/2 TC Haplotype frequency	No of subjects not possessing IGR2096, IGR2198 and IGR2230 variants	OCTN1/2 TC Haplotype frequency
Controls	79	17 (21.5%)	85	23 (27.1%)	66	2 (3%)	63	2 (3.2%)
Crohn's Disease	69	21 (30.4%)	76	13 (17.1%)	62	1 (1.6%)	41	0
P value		0.22		0.13		0.59		0.25

The number of CD patients and controls are shown in the absence of the IBD5 risk haplotype markers IGR2096, IGR2198, IGR2230 and all three SNPs combined. The location of these markers SNPs within the IBD5 locus is illustrated in Figure 6.1. The frequency of the OCTN1/2 TC haplotype within each of these respective groups is illustrated together with P values. In individuals not possessing allelic variants associated with the extended IBD5 haplotype, the OCTN1/2 TC haplotype was not significantly associated with Crohn's disease, for any of the three markers, providing evidence against an independent effect of the OCTN1/2 haplotype on disease susceptibility. In Peltekova's data set the single marker- IGR2078 (block 4) was used to define the IBD5 risk haplotype. It is noteworthy that IGR2096, the only marker of the three studied in our population to show even a trend towards independent segregation from OCTN1/2, was the marker farthest from OCTN1/2, and also was closest to the single marker used by Peltekova et al.

6.3.5 Expression of Genes within the IBD5 Locus

Using the microarry expression data, in Crohn's disease small but consistent downregulation of OCTN2 was observed when Crohn's disease biopsies were compared to control biopsies (FC -1.19, p =0.0028) and inflamed Crohn's disease sigmoid colon biopsies were compared to non-inflamed biopsies (FC -1.46, p =0.00086) (Table 6.7). When Crohn's disease biopsies from the terminal ileal were compared to control terminal ileal biopsies OCTN2 was again downregulated (FC -1.58, p =0.068). Expression of IRF1 was upregulated in the inflamed Crohn's disease sigmoid colon biopsies when compared to the non-inflamed Crohn's disease sigmoid colon biopsies (FC +1.57, p =0.00052).

OCTN2 was downregulated in ulcerative colitis biopsies compared to controls (FC-1.26 p= 3.37×10^{-6}) and when inflamed ulcerative colitis sigmoid colon biopsies were compared to non-inflamed ulcerative colitis sigmoid colon biopsies (FC-1.50, p= 2.2×10^{-6}) (Table 6.8).

Table 6.7: Fold expression changes in genes within the IBD5 locus in patients with Crohn's disease (CD) and controls.

Genes Analyzed	All CD (99) v controls (73) Fold change (FC)	p value	Inflamed CD sigmoid (16) v non-inflamed CD sigmoid (17) (FC)	p value	Non-inflamed CD sigmoid (17) v non-inflamed control sigmoid (18) (FC)	p value	CD TI (16) v HC TI (6) (FC)	p value
IL4	-1.20	0.024	-1.45	0.016	+1.24	0.24	+1.60	0.43
IL13	+1.065	0.21	+1.016	0.87	-1.038	0.71	-1.17	0.75
RAD50	-1.012	0.47	-1.005	0.88	-1.048	0.24	+1.033	0.57
IL5	-1.005	0.91	-1.030	0.66	+1.05	0.45	+1.081	0.56
IRF1	+1.15	$1.7x10^{-13}$	+1.57	0.00052	+1.095	0.0018	+1.17	0.00002
OCTN2 (SLC22A5)	-1.19	0.0028	-1.46	0.00086	+1.31	0.00005	-1.58	0.0068
OCTN1 (SLC22A4)	+1.056	0.60	+1.08	0.64	+1.10	0.40	+1.024	0.61
PDLIM4	+1.02	0.32	-1.036	0.49	-1.075	0.023	-1.12	0.62
P4HA2	-1.025	0.20	-1.040	0.72	+1.36	0.021	+1.045	0.41
CSF2	+1.27	0.00011	-1.022	0.89	+1.12	0.46	+1.19	0.58
IL3	-1.003	0.87	-1.061	0.14	-1.023	0.54	-1.062	0.38

Fold changes in expression of genes within the IBD5 locus comparing controls, patients with inflammatory bowel disease and patients with Crohn's disease. Biopsies from these patients have been stratified for anatomical location, the degree of inflammation and for disease behaviour. TI- terminal ileum. Gene annotation- IL- interleukin, RAD50- DNA repair protein RAD50, IRF1- interferon regulatory factor 1, SLC22- solute carrier family 22, PDLIM4- PDZ and LIM domain 4, P4HA2- Prolyl 4-hydroxylase subunit alpha-2 precursor, CSF2- colony stimulating factor 2.

Table 6.8: Fold expression changes in genes within the IBD5 locus in patients with ulcerative colitis (UC) and controls

Genes Analyzed	All UC (129) v controls (73) (FC)	p value	Inflamed UC sigmoid (35) v non-inflamed UC sigmoid (22) (FC)	p value	Non-inflamed UC sigmoid (22) v non-inflamed control sigmoid (18) (FC)	p value
IL4	+1.00	0.96	+1.06	0.62	+1.14	0.40
IL13	+1.12	0.057	-1.02	0.82	1.00	0.98
RAD50	+1.02	0.20	-1.20	$3.5x10^{-6}$	+1.08	0.062
IL5	+1.09	0.037	+1.10	0.17	+1.04	0.53
IRF1	+1.10	0.35	+1.12	$2.9x10^{-6}$	-1.01	0.62
OCTN2 (SLC22A5)	-1.26	$3.37x10^{-6}$	-1.50	$2.2x10^{-6}$	+1.02	0.75
OCTN1 (SLC22A4)	-1.18	0.11	-1.79	$1.53x10^{-9}$	+1.22	0.63
PDLIM4	+1.10	0.0056	+1.14	0.00039	+1.01	0.78
P4HA2	-1.05	0.13	1.00	0.98	-1.05	0.28
CSF2	+1.10	0.056	+1.19	0.052	-1.04	0.63
IL3	-1.01	0.39	+1.02	0.67	+1.01	0.75

Fold changes in expression of genes within the IBD5 locus comparing controls and patients with ulcerative colitis, who have been stratified for the degree of inflammation observed in their sigmoid colon biopsies.

PHENOTYPIC ANALYSIS

6.3.6 Age of diagnosis

When allelic frequency was compared in sub-groups defined by the Vienna classification for age in the Crohn's disease patients (A1 <40), (A2 >40) there was no significant difference between variant and wild type allelic frequency for all studied SNPs.

6.3.7 Anatomical Distribution

There was no association between the IBD5 marker SNPs, OCTN1/2 variants and Crohn's disease when disease location was assessed by the Vienna classification (L1 terminal ileum, L2 colon, L3 ileo-colonic and L4 upper GI tract). Of note no association was found between the presence of peri-anal Crohn's disease and homozygosity for the OCTN1/2 TC haplotype (p= 0.875). When patients with Crohn's disease were further stratified for NOD2/CARD15 variant carriage, no associations between homozygosity for the OCTN1/2 TC haplotype and the disease location, categorized by the Vienna classification, were observed- 45% of OCTN1/2 TC homozygous Crohn's disease patients with terminal ileal disease (L1) carried no NOD2/CARD15 variants, whereas 52% carried one or more NOD2/CARD15 variant, P=0.75. The IBD5, OCTN1/2 variants were not associated with disease extent and severity in the ulcerative colitis cohort.

6.3.8 Disease Behaviour

When disease behaviour at diagnosis was analyzed, no significant association between variants representing the extended IBD5 locus, individual OCTN1/2 variants and stricturing (B2, Vienna classification) or penetrating (B3) behaviour was observed, when compared to inflammatory, nonstricturing, nonpenetrating disease behaviour (B1). A significant association was observed between IGR2198 homozygosity and stricturing/penetrating disease when compared with inflammatory disease at most recent follow up (24.9% B2, B3 v 14.4% B1, p= 0.026, RR=1.97) (Table 6.9). Significant associations were also observed between stricturing/penetrating disease when compared to inflammatory disease for OCTN1 variant homozygosity (30% B2, B3 v 17.4% B1, p= 0.011, RR=2.0), OCTN2 variant homozygosity (34.8% B2, B3 v 23.0% B1, p= 0.029, RR=1.78) and homozygosity for the OCTN1/2 TC haplotype (29.7% B2, B3 v 17.1% B1, p= 0.011, RR=2.05). In individuals who lacked the IBD5 risk haplotype (homozygous with respect to the non- Crohn's disease associated allele of the IGR2198 variant), there was no association between stricturing/penetrating disease behaviour at follow up and homozygosity for the OCTN1/2 TC haplotype (11.1% inflammatory v 14.9% stricturing/penetrating, p= 0.65).

Table 6.9: Phenotypic associations of IGR2198, OCTN1/2 variant homozygosity and homozygosity for the OCTN1/2 TC haplotype.

Phenotype	% CD Patients homozygous for the variant allele/ haplotype			
	IGR2198	OCTN2 (-207 G→C)	OCTN1 (1672C→T)	OCTN1/2 TC Haplotype
Vienna classification of disease behaviour at follow up				
Inflammatory (n=121)	14.4%	23.0%	17.4%	17.1%
Stricturing/Penetrating (n=217)	24.9%	34.8%	30.0%	29.7%
P Value (RR)	0.026 (1.97)	0.029 (1.78)	0.011 (2.0)	0.011 (2.05)
Disease progression from diagnosis to follow up				
No progression (n=126)	14.7%	23.4%	18.2%	17.8%
Progression (n=132)	25.9%	31.3%	26.9%	29.1%
P value (RR)	0.038 (1.72)	0.18 (1.49)	0.12 (1.66)	0.049 (1.63)
Surgery for luminal complications of Crohn's disease				
No surgery (n=137)	13.1%	21.9%	11.7%	12.3%
Surgery (n=237)	23.2%	33.1%	26.6%	26.7%
P value (RR)	0.037 (1.91)	0.031 (1.77)	0.0007 (2.7)	0.0023 (2.2)

An association was observed between IGR2198 variant homozygosity, OCTN2 variant homozygosity, OCTN1 variant homozygosity and homozygosity for the OCTN1/2 TC haplotype and stricturing/penetrating disease when compared with inflammatory disease at most recent follow up. Of the 258 CD patients who had inflammatory Vienna disease classification at diagnosis, 126 patients disease did not progress at follow up and 132 patients progressed to the stricturing/penetrating group. The average duration of follow up of these patients was 11.8 years.

6.3.9 Disease Progression

As the Vienna classification is a hierarchical system, a number of patients will move from inflammatory disease (B1) at diagnosis to either stricturing (B2) or penetrating disease (B3) during their follow up. Of the 258 patients who had inflammatory disease at diagnosis, 126 patients remained in the inflammatory group at follow up and 132 patients progressed to the stricturing group or the penetrating group. The median duration of follow up of these patients was 11.8 years. Homozygosity rates of IGR2198 variants and the OCTN1/2 TC haplotype were associated with disease progression to Vienna (B2) or Vienna (B3) compared with those whose disease remained as B1 (Table 6.9). The association between the OCTN1/2 TC haplotype and disease progression was not observed in Crohn's disease patients who lacked the IBD5 risk haplotype (18% progression v 13% non-progression, p= 0.56).

6.3.10 Requirement for Surgery

As a marker of disease severity, patients who had required surgery (n= 237) for complications of luminal Crohn's disease were compared to those who had had no surgery for Crohn's disease (n= 137). A significant association was observed between requirement for surgery and homozygosity for variants of IGR2198, OCTN1, OCTN2 and the OCTN1/2 TC haplotype (Table 6.9).

Multiple logistic regression analysis was applied, using a model that considered age at diagnosis, disease behaviour, smoking status, family history, NOD2/CARD15 carrier status and IGR2198 homozygosity or the OCTN1/2 TC haplotype, with the outcome

being surgery for luminal complications of Crohn's disease. Co-linearity between IGR2198 and OCTN1/2 TC homozygosity was evident, and allowed for in modelling. IGR2198 homozygosity (p= 0.016, OR=2.86, CI=1.21-6.76) and the OCTN1/2 TC haplotype (p= 0.004, OR=3.52, CI=1.49-8.31) were significantly associated with the need for surgery.

6.4 DISCUSSION

The present study has provided novel data regarding the contribution of the IBD5 locus as a determinant of disease susceptibility and phenotype in the Scottish population, which is known to be characterized by low rates of racial admixture, compared with others in Europe, and North America.(251) We have shown the OCTN1 and OCTN2 polymorphisms to be in tight linkage disequilibrium (D′ >0.959) with allelic variants of the 3 SNPs representing the extended IBD5 haplotype (IGR2096, IGR2198 and IGR2230). Susceptibility to Crohn's disease was associated with each of the three IBD5 polymorphisms defining the extended IBD5 region, as well as the variant alleles of the OCTN1 and OCTN2 genes. Homozygosity for each of the three SNPs used to define the extended IBD5 haplotype and for the OCTN1/2 TC haplotype were strongly associated with disease susceptibility. These data provide the first independent confirmation of the association with the OCTN1/2 variants studied by Peltekova and colleagues in Canada.(95)

However, our data are not entirely consistent with the results reported from Peltekova and colleagues, in a potentially critical aspect. In the absence of allelic variants representing the extended risk haplotype for IBD5, our results show no significant difference in the presence of the OCTN1/2 TC haplotype in Crohn's disease patients compared to controls. These findings lead to the important suggestion that the OCTN1/2 TC haplotype may not confer risk of Crohn's disease independently of other closely linked determinants within the extended IBD5 locus, and that it is at present premature to conclude that the OCTN1/2 variants are causative in the pathogenesis of Crohn's disease.

In comparing the studies reported from Peltekova and colleagues with our own, the choice of markers used to define the IBD5 locus is especially worthy of discussion. In Peltekova's data set the single marker IGR2078, located in block 4 of the haplotype map (89) was used to define the IBD5 risk haplotype, and not a marker closer to the OCTN1/2 loci (block 7), such as IGR2198 (block 5) or IGR2230 (block 7) which were used in the present study. In light of this, recombination between the IGR2078 marker in block 4 and OCTN1/2 in block 7 needs to be considered as an explanation for the apparent independence reported by Peltokova and colleagues. In our data the only marker for which any trend towards an independence of the OCTN1/2 TC haplotype was observed concerned the IGR2096 variants, which is also located in block 4. In the absence of variants of the markers IGR2198 and IGR2230, the OCTN1/2 TC haplotype was in fact more prevalent in the control group, further supporting the hypothesis of recombination between haplotype blocks four and seven.(89)

Data from four large cohorts of inflammatory bowel disease patients in Cambridge,(99) Stockholm, Sweden,(252) UK/ Germany (100) and the USA (253) have all shown an association between the IBD5 locus and Crohn's disease. In each of these data sets, as in the present study, there was no independent association between the OCTN1/2 TC haplotype and Crohn's disease in the absence of the extended IBD5 risk haplotype. The association between the IBD5 locus and Crohn's disease has also been convincingly replicated in the recent meta-analysis of genome wide association studies with a Bonferroni corrected p value of less than 0.05.(151)

The arguments as to the contribution of the OCTN1/2 variants, implicated by Peltekova and colleagues, have been developed further, and clearly require resolution. Studies involving 276 samples from healthy, ethnically diverse human populations have allowed the identification of several non-conservative SNPs of OCTN1/2 in evolutionary conserved sites in African-American and Chinese populations.(254) The authors suggest that if these variants were shown to confer susceptibility to Crohn's disease in their respective populations this would provide strong evidence that OCTN1/2 contain the critical mutations. As yet these studies have not been undertaken. Moreover, genetic mutations in OCTN2 cause systemic carnitine deficiency, characterised by disease of skeletal muscle, cardiac muscle and liver, but not inflammatory or intestinal disease.(255)

Genetic studies to resolve the limits of the association using markers p and q telomeric to OCTN1/2 may help to clarify this issue, however, power of resolution will become a critical issue in attempting to resolve this controversy by genetic studies alone. We have calculated that 3200 Crohn's disease individuals would need to be genotyped to prove the independence of OCTN1/2 from IGR2230 which is situated between the two genes in haplotype block 7.(89)

In the largest study of IBD5 to date, Silverberg and colleagues used stepwise regression analysis to investigate 1879 patients with inflammatory bowel disease. The effect of three SNPs IGR2096, IGR2198 and OCTN1 (1672C→T) were individually enough to explain the effect of the IBD5 locus in Crohn's disease susceptibility.(253) OCTN2 (-207 G→C) was ruled out as the causal variant, however, the study did not

have sufficient power to differentiate between the 3 remaining SNPs and a meta-analysis of the published data may help define the critical variant.

When gene expression was compared within the IBD5 locus consistent downregulation of OCTN2 expression was observed in Crohn's disease and ulcerative colitis biopsies compared to controls. In Crohn's disease it was the inflamed biopsies that were driving these changes and OCTN1 was also downregulated in inflamed ulcerative colitis biopsies. These provocative data add weight to the hypothesis that decreased expression of the OCTN1/2 genes may be involved in the pathogenesis of IBD, however, we also noted expression of IRF-1- a transcriptional factor that is involved in the regulation of a number of genes related to the innate and adaptive immune response (256;257) to be dysregulated, emphasizing the uncertainties pertaining to this locus at present.

Increasingly in Crohn's disease it is recognised that clinical phenotype is genetically determined.(258) In our cohort, when the Vienna classification for disease behaviour was analyzed at the patients' most recent follow up assessment, there was a significant association between the IBD5 marker SNP IGR2198, OCTN1/2 variants, the OCTN1/2 TC haplotype and the presence of stricturing and penetrating disease. Crohn's disease patients who were homozygous for IGR2198 variants or homozygous for the OCTN1/2 TC haplotype had a disease phenotype that was more likely to progress to stricturing or penetrating disease behaviour, and multivariate analysis showed an association between IGR2198, the OCTN1/2 TC haplotype and the requirement for surgery.

Further more recent data from a cohort of 325 German patients using the IGR2096 SNP to represent the IBD5 haplotpye showed variant alleles were associated with stricturing/ fistulising disease behaviour compared to inflammatory disease behaviour and the effect was dose related with homozygous patients having a higher risk (OR 6.5) than heterozygous patients (OR 4.8).(259) Variants within the IBD5 locus have also been associated with a more severe disease phenotype and growth failure in an independent cohort of 200 Scottish patients, diagnosed with Crohn's disease at the age of sixteen or under.(98) Overall, our data show an exciting, novel genotype-phenotype association with the IBD5 locus and it is intriguing to propose that a genetic variant within the IBD5 locus causes Crohn's disease to become more severe (stricturing or penetrating disease) and hence require increased surgical intervention.

Previous work involving the IBD5 locus has shown an association between the IBD5 risk haplotype and peri-anal Crohn's disease but this was not replicated in our cohort. (93) Phenotypic data from Newman et al showed an association with ileal disease (112), however, no significant association was observed in our population. Differences in definitions of ileal Crohn's disease may be the reason for the differences between these two cohorts: ileal disease in our cohort was classified strictly according to the Vienna classification (L1). Using these criteria, 34% of patients had purely terminal ileal disease: in contrast, 70.6% of patients in the Toronto cohort were classified as having ileal disease.(112) Newman and colleagues did not appear to include any patients with proximal small intestinal or upper gastrointestinal disease. No data were available in this cohort regarding disease behaviour, or disease progression over time.

In conclusion we have determined that the IBD5 locus is associated with disease susceptibility in our Crohn's disease cohort in Scotland. This is the first independent replication of the association of the OCTN1/2 haplotype with Crohn's disease outwith the index population in Canada. However, a significant effect of OCTN1 and OCTN2 variants was not seen in the absence of the IBD5 risk haplotype, and on the strength of the present data it is not possible to conclude whether or not the OCTN1/2 genes contain the disease-causing mutation, or whether the association serves to only narrow the region of association within the IBD5 locus. Finally, novel and potentially important phenotypic associations have been identified with the IBD5 locus, which implicate this region as a determinant of disease severity in Crohn's disease, as well as susceptibility.

CHAPTER 7

DLG5 VARIANTS DO NOT INFLUENCE SUSCEPTIBILITY TO INFLAMMATORY BOWEL DISEASE IN THE SCOTTISH POPULATION

SUMMARY

Introduction and Aims: Data from Germany have suggested that specific haplotypic variants of the DLG5 gene on chromosome 10q23, may be associated with susceptibility to inflammatory bowel disease (IBD). Association of DLG5 haplotypic variants with disease susceptibility, genotype-phenotype relationships and epistasis with CARD15 was investigated in the Scottish population. Expression of DLG5 was also investigated in human colonic and ileal endoscopic biopsies.

Patients and methods: 374 CD, 305 ulcerative colitis (UC) and 294 healthy controls (HC) were studied. Genotyping for the variants rs1248696 (113A, representing haplotype D) and the SNP tag rs2289311 (representing haplotype A) were typed using the Taqman system. Expression of the DLG5 probe A_23_P161209 on the *Agilent* microarray chip was investigated in 53 CD, 67 UC and 31 control subjects.

Results: On analysis of the DLG5 variant 113A there were no associations with IBD when allelic frequency (11.4%IBD v 13.2%HC, P=0.30) and carrier frequency (19.2%IBD v 24.6%HC, p= 0.069) were analyzed. No associations were observed between 113A variant allelic frequency (p= 0.37), carrier frequency (p= 0.057) and CD. No associations between DLG5 and UC were observed. Haplotype A was not protective and there was no evidence of epistasis between DLG5 and CARD15. No significant changes in expression of DLG5 were observed in any of the disease groups that were examined.

Conclusions: The present data contrast strongly with previous data from Germany. DLG5 113A is not associated with disease susceptibility and haplotype A does not confer resistance.

7.1 INTRODUCTION

A locus on chromosome 10 was first implicated in genome wide scanning of a European cohort (United Kingdom, Germany and the Netherlands) in 1999 as conferring susceptibility to inflammatory bowel disease (LOD score 2.30).(78) Results now suggest that the gene DLG5 (Drosophila Discs Large Homolog 5) located on chromosome 10q23 may be responsible for the observed linkage and contain critical mutations that confer susceptibility to inflammatory bowel disease.(107)

Stoll et al (107) identified two extended DLG5 haplotypes that influenced disease susceptibility in the German population. The first haplotype, (named haplotype D) was especially notable for the presence of a G→A substitution at nucleotide 113 that resulted in an amino change at position 30 from Arginine to Glutamine (R30Q). On analysis of carrier frequency, Stoll found the 113A variant to be associated with Crohn's disease (25% Crohn's disease v 17% healthy controls, P=0.001) in a case control study and trends between 113A transmission and inflammatory bowel disease (p= 0.09) and Crohn's disease (p= 0.065) were observed on transmission disequilibrium testing.(107) *In silico* analysis suggests that the 113A (R30Q) variant may impair DLG5 scaffolding function, but as yet no expression or functional studies in the inflammatory bowel diseases have been conducted. Evidence of epistasis between the 113A variant of DLG5 and CARD15 variants was also observed in the Crohn's disease cohort.(107)

The second haplotype, (haplotype A) was tagged by eight marker SNPs and was observed to be significantly under-transmitted in the inflammatory bowel disease group (p= 0.006) suggesting the haplotype may be protective.(107)

In the present study we have assessed the contribution of the DLG5 polymorphisms rs1248696 (113A) and rs2289311 (one of the marker SNPs for the protective haplotype A)(107)) in determining genetic susceptibility to Crohn's disease and ulcerative colitis in the Scottish population which has a high incidence of inflammatory bowel disease. We have also investigated genotype-phenotype associations in our rigorously defined inflammatory bowel disease population and examined expression of DLG5 in the colon and terminal ileum of patients with inflammatory bowel disease and controls.

7.2 PATIENTS AND METHODS

7.2.1 Demographics: Crohn's Disease and Ulcerative Colitis.

Six hundred and seventy nine patients with well characterized inflammatory bowel disease and 269 controls were recruited. The group comprised of 374 patients with Crohn's disease and 305 patients with ulcerative colitis (Table 7.1) (Methods 3.4.1-3). The duration of follow-up was defined as the time from diagnosis to the time of most recent clinic review (median duration of 11.8 years in the Crohn's disease group and 7.5 years in the ulcerative colitis group). The Crohn's disease group consisted of 181 males and 193 females and the ulcerative colitis group consisted of 171 males and 134 females with a median age at diagnosis of 34 years. Vienna disease classification was available for 347 (93%) of Crohn's disease patients at diagnosis and 374 (100%) of Crohn's disease patients at follow up. Full phenotypic data were available for the ulcerative colitis cohort.

7.2.2 Control Subjects

Two hundred and ninety four controls- 163 blood donors from the south east of Scotland and 131 healthy controls subjects were enrolled. Allelic frequencies of DLG5 variant single nucleotide polymorphisms (SNPs) 113A, rs2289311, OCTN1 variant (SLC22A4 1672C→T) and OCTN2 variant (SLC22A5 -207G→C) are shown in Table 7.2.

Table 7.1: Demographics and clinical features of the Crohn's disease, ulcerative colitis and control group.

	Crohn's Disease	Ulcerative Colitis	Controls
Total Number	**374**	**305**	**269**
Sex (male / female)	181/193	171/134	135/134
Median Age at diagnosis (yrs) Interquartile range	27.8 20.9-40.4	34 25-50	39 27-52
Median duration of follow up (yrs) Interquartile range	11.8 6.5-20.2	7.5 3.35-13.38	
Caucasian (%)	98.7%	98.3%	
CD Location at Diagnosis		**UC disease extent**	
Ileal disease	125 (36%)	Proctitis 105 (34.5%)	
Colonic Disease	137 (39%)	Left sided colitis 116 (37%)	
Ileal and colonic	58 (14%)	Extensive colitis 84 (27.5%)	
Upper GI disease	30 (8.5%)		
Perianal disease	75 (21.4%)		
CD Location at Follow Up			
Ileal disease	92 (25%)		
Colonic Disease	130 (35%)		
Ileal and colonic	103 (28%)		
Upper GI disease	49 (13%)		
CD Behaviour at Diagnosis			
Inflammatory (Vienna B1)	258 (74.8%)		
Stricturing (Vienna B2)	30 (8.4%)		
Penetrating (Vienna B3)	59 (16.8%)		
CD Behaviour at Follow Up			
Inflammatory (Vienna B1)	142 (38%)		
Stricturing (Vienna B2)	68 (18%)		
Penetrating (Vienna B3)	164 (44%)		

Table 7.2: Demographics and allelic frequencies in the blood transfusion samples and healthy volunteer control samples.

	Blood Transfusion Samples	Healthy Control Samples
Age Median (IQ range)	35 (26-47)	36 (29-51)
Male/ Female	83/ 79	52/ 55
113A (Allelic Frequency)	14.4%	12%
rs2289311	30.4%	33%
OCTN1 (SLC22A4)	43.5%	42.4%
OCTN2 (SLC22A5)	49%	46.1%
IGR2198	42%	40.4%

Allelic frequencies of DLG5 variant SNPs 113A, rs2289311, OCTN1 variant (C→T), OCTN2 variant (-207G→C) and IBD5 marker SNP IGR2198 are shown to illustrate the consistency between the blood transfusion controls and the healthy volunteer controls.

7.2.3 Genotyping

The SNPs rs1248696, (113G→A representing haplotype D) and rs2289311 [(chosen because of its reliability in genotyping to represent the protective haplotype A), *M Stoll, personal communication*] were typed using the Taqman system (Methods 3.4.4-6). Inflammatory bowel disease patients and controls were typed for polymorphisms of the CARD15 gene (R702W, G908R, and 1007fsinsC) using previously described methods.(85) All genotyping except the R702W was carried out using the Taqman system. R702W genotyping was performed by restriction fragment length polymorphism polymerase chain reaction.

7.2.4 Data Analysis

The two SNPs rs1248696 and rs2289311, were analyzed for association with Inflammatory bowel disease (IBD) overall, Crohn's disease, ulcerative colitis and disease phenotype. Each allele was shown to be in Hardy-Weinberg equilibrium in the

control population. Genotype-phenotype associations were analyzed by Chi squared test using Minitab v10. To identify significant independent variables associated with genotype, univariate and multivariate analysis was carried out. Evidence for DLG5 epistasis with CARD15 was investigated by stratifying DLG5 variants by carriage of one or more of the three common CARD15 variants- R702W, G908R, and 1007fsinsC. Allelic frequencies of the DLG5 variants were compared between the sub-groups of patients with and without CARD15 variants by Chi squared analysis. The null hypothesis was that the frequency of DLG5 variants did not differ between these sub-groups. Phenotypic associations of DLG5 variants were also stratified for the presence and absence of CARD15 variants.

7.2.5 Expression of the DLG5 probe A_23_P161209

Gene expression of DLG5 was analyzed using data generated from the microarray gene expression study. Detailed methods are provided in the methods section 3.2.1-9, 3.3.4. Statistical significance of the microarray data was determined by Student's unpaired *t* test. $p < 0.01$ and a fold change of greater or less than 1.5 were considered statistically significant.

7.3 RESULTS

7.3.1 Disease Susceptibility: Haplotype D (113A)

On analysis of the DLG5 variant 113A there were no associations with inflammatory bowel disease (IBD) when allelic frequency (11.4% IBD v 13.2% HC, p = 0.30), carrier frequency (19.2% IBD v 24.6% HC, p = 0.069) and homozygosity rates (2.3% IBD v 1.5% HC, p = 0.48) were analyzed (Table 7.3). A negative association was observed between heterozygous rates of 113A and IBD (16.9% IBD v 23% HC, p = 0.033). Furthermore a negative correlation was observed between heterozygous 113A variants and Crohn's disease (CD) (16% CD v 23% HC, p = 0.029). No association was observed between 113A variant allelic frequency (11.4% CD v 13.2% HC, p = 0.37), carrier frequency (18.3% CD v 24.6% HC, p = 0.057), homozygous rates (p = 0.36) and CD. No associations between 113A and UC were observed-allelic frequency (12.8% UC v 13.2% HC, p = 0.34), carrier frequency (20.3% UC v 24.6% HC, P=0.23) heterozygous rates (18% UC v 23% HC, p = 0.45) and homozygous rates (2.3% UC v 1.5% CD, p = 0.5).

Table 7.3: DGL5 113A variant allele frequency, carrier frequency, heterozygote frequency and homozygote frequency in the IBD, CD, UC and control populations.

	Controls	IBD p value	Crohn's Disease p value	Ulcerative Colitis, p value
Allelic Frequency	13.2%	11.4% p=0.30	11.4% p=0.37	11.4% p=0.34
Carrier Frequency	63/256 (24.6%)	125/652 (19.2%) p=0.069	65/356 (18.3%) p=0.057	60/296 (20.3%) p=0.228
Heterozygosity Rates	59/256 (23%)	110/650 (16.7%) **p=0.033**	57/356 (16%) **p=0.029**	53/294 (18%) p=0.45
Homozygosity Rates	4/256 (1.5%)	15/650 (2.3%) p=0.48	8/356 (2.2%) p=0.36	7/294 (2.4%) p=0.50

The p values shown are calculated between the control group and each respective disease group.

7.3.2 DLG5 Haplotype A

Haplotype A allelic frequencies, represented by rs2289311 variants did not differ between HC (31.5%) and IBD (35%, P=0.17), CD (36.9%, P=0.078) or UC patients (33.4%, P=0.51). No significant differences were observed between carriage rates of rs2289311 variants, HC (52%), IBD (57.2%, P=0.18), CD (60.9%, P=0.052) and UC (54%, P=0.65). The frequency of patients who were heterozygote or homozygote for rs2289311 polymorphisms did not differ between IBD and HC groups (heterozygote HC 41.1% v IBD 44.8% P=0.43, CD 48% P=0.13 and UC 42.1%, P=0.83) (homozygote HC 10.9% v IBD 12.6% P=0.49, CD 12.6% P=0.61 and UC 12.9%, P=0.50).

7.3.3 Phenotypic Analysis

On univariate analysis of Crohn's disease patients no association was observed between DLG5 113A variants and the Vienna classification for age of diagnosis, location of disease and disease behaviour. The location of disease and disease behaviour in Crohn's disease patients was analyzed at time of diagnosis and at most recent follow up and there was no association between DLG5 113A variants and disease progression. No association was observed between DLG5 113A variants and age of diagnosis in the inflammatory bowel disease and ulcerative colitis groups and there was no association between DLG5 113A variants, disease extent and severity in the ulcerative colitis patients.

When the frequency of the DLG5 113A was compared in males and females, there was no difference in both the control, inflammatory bowel disease group and Crohn's disease group– control 12.4% male v 12.8% female, p = 0.79, IBD male 12.2% v female 9.5%, p = 0.14, Crohn's disease, male 9.9% v female 10.5%, p = 0.68. In the ulcerative colitis group the DLG5 113A allele was more common in males 14.3% than females 7.9%, p = 0.022, OR 1.9, CI (1.1- 3.3).

DLG5 113A variants displayed a trend towards being less common in inflammatory bowel disease patients with joint problems (large joint arthralgias related to disease activity, small joint arthralgias unrelated to disease activity, ankylosing spondylitis and sacro-ilitis) (N=127) when compared to those who had no joint problems when allelic frequency was analyzed (7.5% v 11.8%, p = 0.053). When allelic frequency of 113A variants was analyzed in ulcerative colitis patients with primary sclerosing

cholangitis a trend towards these patients having fewer 113A variants was observed (0%, (N=7) v 11.4%, p = 0.17). On analysis of the haplotype A, no genotype phenotype associations were observed in the Crohn's disease and ulcerative colitis patient groups. Multiple logistic regression analysis did not identify any variables that were independently associated with haplotype D (113A) or haplotype A.

There was no evidence of epistasis between DLG5 113A variants and carriage of the three common CARD15 variants Gly908Arg, Arg702Trp and Leu1007fsinsC-CARD15 carriage positive DLG5 113A allelic frequency 9.7% (N=108) v CARD15 carriage negative DLG5 113A allelic frequency 11.6% (N=584) (p = 0.43). When Crohn's disease patients were stratified for CARD15 variant carriage, no significant genotype-phenotype relationships were found with DLG5 113A.

7.3.4 Expression of DLG5

Using the Agilent probe A_23_P161209 to represent DLG5, expression of the gene was compared across a number of experiments (Table 7.4). No significant changes in expression were observed when inflammatory bowel disease, Crohn's disease and ulcerative colitis biopsies were compared to control biopsies (p= 0.78, p= 0.12 and p= 0.063 respectively). When inflamed and non-inflamed Crohn's disease and ulcerative colitis sigmoid colon biopsies were compared, again no change in expression was observed between the inflamed and non-inflamed groups (p= 0.29 and p= 0.94 respectively).

Table 7.4: Expression of DLG5 probe A_23_P161209 across a number of different experiments

Groups Analyzed (Number of biopsies)	Fold Change	p value
All IBD (228) v controls (73)	-1.0057	0.78
All CD (99) v controls (73)	+1.082	0.12
All UC (129) v controls (73)	-1.046	0.063
Non-inflamed CD sigmoid (17) v non inflamed control sigmoid (18)	-1.044	0.55
Non-inflamed UC (22) v non-inflamed control sigmoid (18)	-1.12	0.53

7.4 DISCUSSION

The present study has demonstrated that in the Scottish population, traditionally characterised by low rates of admixture, the DLG5 variant 113A representing haplotype D is not a critical determinant of susceptibility in either Crohn's disease or ulcerative colitis. Haplotype A represented by the SNP rs2289311 was not protective in our Crohn's disease or ulcerative colitis population and no change in expression of DLG5 was observed in any of the disease groups that were examined.

These data differ markedly from those of Stoll et al who showed that in a German population, DLG5 113A variants were overtransmitted to individuals with inflammatory bowel disease, and in a case control study there was significantly higher rates of 113A carriage in the inflammatory bowel disease group when compared to the control group (25% v 17%, p = 0.001).(107) Considerations including sample size and phenotypic differences between the present study and that of Stoll and colleagues may be responsible for the observed discrepancy in results, but a more plausible explanation would be genetic heterogeneity between the populations of Germany and Scotland. This has been clearly illustrated by data now available with respect to the three common polymorphisms of the NOD2/ CARD15 gene (G908R, R702W and 100fsinsC) which are significantly more common in the Central European Crohn's disease population (42;82) than in the Northern European Crohn's disease population.(85;260;261) Furthermore CARD15 polymorphisms are absent in the Japanese and Chinese Crohn's disease populations.(83;84;262) The different incidences of the R702W polymorphism has also been shown in healthy volunteers in Europe, Africa and Asia.(263) At

present DLG5 has not been associated with inflammatory bowel disease in any genome wide association study and this would suggest that if indeed DLG5 plays a role in the pathogenesis of inflammatory bowel disease its contribution may be small and limited to specific populations.(151;152;264)

Further data illustrating genetic heterogeneity in European inflammatory bowel disease patients have been observed on analysis of the Asp299Gly mutation of the Toll-like receptor 4 gene (TLR4).(265) TLR4 is a member of the Toll like receptor family which are involved in recognition of pathogen associated molecular patterns by the immune system and TLR4 functions as an extracellular pattern recognition receptor for lipopolysaccaride which is common to many intraluminal bacteria.(266) The Asp229Gly variant has been shown to confer susceptibility to Crohn's disease and ulcerative colitis in the Belgian population, (267) but no association was observed between Asp229Gly variants and inflammatory bowel disease in German and Scottish cohorts.(85;268)

Analysis of the original German cohort of patients used by Stoll and colleagues,(107) combined with the Italian and Canadian patients used by Daly and colleagues,(120;135) showed a significant association between the 113A allele and Crohn's disease males, $p < 0.001$, OR = 2.49, but not females, $p = 0.979$, OR = 1.01.(135) The observed association of the 113A variant and Crohn's disease in males appeared to be driven by the difference between the allelic frequencies of the 113A DLG5 variant in the control male (5.2%) compared to the female population (11.3%).

A meta-analysis of 12 case-control studies including the present data containing 4707 Crohn's disease patients and 4937 controls showed no male-female allelic frequency differences in the control populations and a trend towards the 113A/R30Q variant being less common in the female Crohn's disease patients.(136) When these data were added to further analysis of 5 previous case-control studies the 113A/R30Q variant was associated with a small reduction in risk in female Crohn's disease patients- OR = 0.86 (CI 0.76- 0.97).

Interestingly, in a cohort of patients diagnosed with inflammatory bowel disease at the age of 16 or under from Scotland, Russell and colleagues observed an overtransmission of the 113A /R30Q allele in male patients with Crohn's disease compared to females- 29.3% vs. 16.9%, p=0.04 OR 2.04.(269) Taken together these data suggest that if there is a genetic contribution of the DLG5 113A/R30Q variant to disease susceptibility it may be more prominent in the male population.

Trends were observed towards a lower DLG5 113A variant frequency in Crohn's disease patients with arthropathy and ulcerative colitis patients with primary sclerosing cholangitis (PSC) on univariate analysis. Both of these extra-intestinal complications have been shown to have a molecular genetic basis. Susceptibility to axial arthropathy has been strongly associated with HLA-B*27 in patients with and without inflammatory bowel disease (270) and peripheral arthropathies in patients with inflammatory bowel disease have been associated with HLA-DRB1*0103, HLA-B*35, HLA-B*27 and HLA-B*44.(271) In PSC strong disease associations with extended HLA haplotypes have been observed (272) and a functional variant of stromolysin (matrix metalloproteinase 3) has also been associated with susceptibility to

PSC and with disease progression.(273) In the current investigation the relatively small numbers of inflammatory bowel disease patients studied with these specific extra intestinal manifestations mean that these results regarding any DLG5 effect in these sub-groups should be regarded as exploratory observations. Replication studies in other cohorts may help shed some light on this question.

No evidence of epistasis between DLG5 113A variants and carriage of the three common NOD2/ CARD15 variants Gly908Arg, Arg702Trp and Leu1007fsinsC was observed in patients with Crohn's disease. Again these data contrast with Stoll and colleagues who found a significantly greater transmission of DLG5 113A in patients with Crohn's disease carrying one of the risk associated alleles of CARD15.(107) A possible explanation for the absence of epistasis between DLG5 and CARD15 could be the low incidence of CARD15 variants in the Scottish Crohn's disease population, 1007fsinsC= 4.7%, G908R=1.8% and R702W= 7.1% and the combined population attributable risk these variants confer 11%.(85)

In conclusion, in our North European study population we were unable to replicate Stoll's data that the DLG5 variant 113A confers susceptibility to inflammatory bowel disease. Haplotype A represented by the SNP rs2289311 does not confer protection in our population. There was no change in expression of DLG5 in the colon and terminal ileum of inflammatory bowel disease patients and controls and this allied to the lack of association in the recent genome wide association studies would strongly suggest that DLG5 does not play a significant role in the pathogenesis of inflammatory bowel disease.

CHAPTER 8

CONCLUSIONS AND FUTURE DIRECTIONS

8.1 CONCLUSIONS

In the healthy adult colon cluster analysis showed differences in gene expression between the right and left colon. (χ^2=25.1, p<0.0001). Developmental genes, HOXA13, ($p=2.3x10^{-16}$), HOXB13 ($p <1x10^{-45}$), GLI1 ($p=4.0x10^{-24}$), and GLI3 ($p=2.1x10^{-28}$) were involved in driving these changes. These novel findings have important implications for disease pathogenesis in the colon.

When all the Crohn's disease biopsies were compared to the controls, 259 sequences were upregulated and 87 sequences were downregulated. Upregulated genes in Crohn's disease included SAA1 (FC +7.5, $p =1.47x10^{-41}$) and REGL (FC +7.3, $p =2.3x10^{-16}$). Downregulated genes included genes involved in cellular detoxification-SLC14A2 (FC -2.49, p =0.00002), carbonic anhydrase 2 (CA2) (FC -2.4, $p =8.4x10^{-10}$) and CA1 (FC -2.3, $p =7.5x10^{-6}$). In the Crohn's disease terminal ileal biopsies diubiquitin (FC+11.3, $p<1x10^{-45}$), MMP3 (FC+7.4, $p=1.3x10^{-11}$), IRTA1 (FC-11.4, $p=4.7x10^{-12}$) and CCL23 (FC-7.1, $p=1.6x10^{-10}$) were differentially expressed compared to controls. In the colon SAA1 (FC+6.3, $p=5.3x10^{-8}$) was upregulated and TSLP (FC-2.3, $p=2.7x10^{-6}$) was downregulated comparing non-inflamed Crohn's disease and control biopsies, and the colonic inflammatory Crohn's disease signature was characterized by downregulated organic solute carriers-SLC38A4, SLC26A2 and OST alpha.

Analysis of the IL-23 pathway revealed that IL-23A/p19, (FC +2.32, p =0.000099), JAK2, (FC +1.90, $p =9.4x10^{-7}$), STAT3, (FC +2.23, p =0.0004) and INFγ, (FC +2.31, p =0.0019)

were significantly upregulated in the Crohn's disease biopsies compared to controls and when inflamed Crohn's disease biopsies were compared to non-inflamed Crohn's disease biopsies- IL-23A/p19, (FC +2.11, p =0.000031), JAK2, (FC +1.90, p =0.00003), STAT3, (FC +1.66, p =0.0002) and INFγ, (FC +2.33, p <0.0001). Modest changes in expression were also observed in a number of the autophagy genes.

When the anatomical location that the Crohn's disease and control biopsies were taken from was considered, 18 terminal ileal biopsies clustered together, separately from the colonic biopsies (6 control and 12 Crohn's disease, $X^2 = 10.2$, $p < 0.001$). No separation of the biopsies by either disease status or by the degree of inflammation was observed. When the terminal ileal biopsies alone were considered, clustering by disease status was observed. Gene ontology showed this clustering was driven by a downregulated group of transporter genes and an upregulated group of structural genes. A panel of 14 epithelial cell cytokines showed good ability to cluster biopsies from Crohn's disease patients and controls emphasising their key role in inflammation.

When all ulcerative colitis biopsies and control biopsies were compared, 143 sequences had a fold change of >1.5 in the ulcerative colitis biopsies ($0.01>p>10^{-45}$) and 54 sequences had a fold change of <-1.5 ($0.01>p>10^{-20}$). Differentially upregulated genes in ulcerative colitis included SAA1 ($p<10^{-45}$) the alpha defensins, DEFA5&6 ($p=0.00003$ and $p=6.95\times10^{-7}$ respectively), MMP3 ($p=5.6\times10^{-10}$) and MMP7 ($p=2.3\times10^{-7}$). When non-inflamed sigmoid colon biopsies were compared between ulcerative colitis patients and controls upregulated genes included defensin beta 14 (FC+2.11, p = 0.00002) and SAA1 (FC+2.01, p = 0.00024).

Interesting genes that were down regulated included HLA-DRB1 (FC-3.0, $p = 0.0010$) and TSLP (FC-2.73, $p = 2.7x10^{-10}$).

In the inflamed ulcerative colitis biopsies there was marked upregulation of DEFA5 and DEFA6 expression in the colon compared to non-inflamed ulcerative colitis biopsies and control biopsies. This was characterized to Paneth cell metaplasia by immunohistochemistry and *in-situ* hybridization. Sub-analysis of the IBD2 loci and the ABC transporter genes revealed a number of differentially regulated genes in ulcerative colitis.

One of the limitations of the expression data set was the lack of genotypic data in the patients and controls. Having DNA in these patients would have allowed a direct comparison between genotype and expression analysis.

In our inflammatory bowel disease population it was observed that the IBD5, OCTN1 and OCTN2 polymorphisms were in strong linkage disequilibrium and variants in all the examined SNPs- IGR2198, IGR2198, OCTN2 (-207 G→C), IGR2230 and OCTN1 (1672C→T) were associated with Crohn's disease compared to healthy controls ($p<0.03$). IGR2198 variants and the OCTN1/2 TC haplotype were associated with a more severe disease phenotype- stricturing/ penetrating disease at follow-up ($p=0.011$, $p=0.011$), disease progression from inflammatory disease phenotype to stricturing/penetrating disease phenotype ($p=0.038$, $p=0.049$), and with need for surgery ($p=0.006$, $p=0.004$ respectively).

In the absence of the IBD5 risk haplotype, no association of OCTN1/2 variants with Crohn's disease was detected. No associations were seen with ulcerative colitis. When the expression data was compared using 11 probes representing genes from the IBD5 locus, small but consistent downregulation of OCTN2 was observed in the Crohn's disease and ulcerative colitis biopsies compared to control biopsies and in inflamed compared to non-inflamed Crohn's disease and ulcerative colitis biopsies.

Analysis of the DLG5 variant 113A showed there were no associations with inflammatory bowel disease, Crohn's disease or ulcerative colitis. DLG5 haplotype A was not protective in our inflammatory bowel disease population. Using the probe A_23_P161209 to represent DLG5, no significant changes in expression were observed when inflammatory bowel disease, Crohn's disease and ulcerative colitis biopsies were compared to control biopsies.

Overall these data emphasise the key role of a number of inflammatory molecules and pathways in the pathogenesis of Crohn's disease and ulcerative colitis, and their potential for translation to therapeutic targets. Furthermore, the results add considerably to the recent genome wide association studies in providing complimentary human colonic and ileal expression data along with detailed analysis of the IL-23 and autophagy pathways.

8.2 FUTURE DIRECTIONS

The expression data presented in this thesis have allowed us to gain insight into how recognized and novel molecules are expressed in the colon and in the terminal ileum in patients with quiescent and active inflammatory bowel disease. One of the next analyses to be undertaken using the existing data set will be to see if small numbers of probes can be identified to correctly assess inflammatory bowel disease in order to make a correct diagnosis and also to grade disease severity. A diagnostic and severity chip may help in the diagnosis of inflammatory bowel disease especially in patients where it is difficult to tell whether they have Crohn's disease or ulcerative colitis.

To further investigate the pathogenesis of inflammatory bowel disease and better define the roles of the novel and established genes investigated in this thesis, the next set of experiments to be undertaken would be functional studies. Mouse models, colonic cell lines and human endoscopic biopsies could be used to examine how central these molecules are to the inflammation observed in inflammatory bowel disease. It would also be interesting to examine the key inflammatory molecules at the protein level and compare this to the RNA expression results. This would help with the ultimate aim of dissecting the immune response in inflammatory bowel disease in order to identify new therapeutic targets.

A further avenue for gene expression studies is to investigate whether they can predict response to therapy and allow clinicians to personalize treatment, maximizing efficacy and

minimizing the risks of therapy. Data has recently been accepted for publication looking at predicting response to infliximab therapy in patients with ulcerative colitis.(274) Gene expression analysis in Crohn's disease before and after therapy especially in terminal ileal biopsies could investigate determinates of response to treatment. Furthermore, it is imperative for new biological therapies to have a diagnostic test to go with the therapy and gene expression could provide this valuable data.

We have also undertaken genome wide expression analysis on whole blood from a number of patients in this study and we also plan to publish this data and compare gene expression in the colon, ileum and peripheral blood. Preliminary results show interesting changes between expression in the bowel and peripheral blood and these results may help in diagnosis and grading severity.

The whole expression data set has also been made available to scientists in Edinburgh, *Genentech* and to the wider scientific community through Gene Expression Omnibus, allowing investigators to look at expression of genes of interest. This has allowed the development of a number of successful collaborations.

With regards to IBD5 we have calculated that 3200 Crohn's disease patients would be required for a study to have the power to identify the disease causing variant and a meta-analysis with sufficient power could be put together with collaborators. At present there are no plans for large scale detailed genotyping of this area, but as genotyping costs continue to fall

this will become increasingly feasible. Subsequent studies into DLG5 have not observed any association with Crohn's disease and we would not propose any further studies into this gene in our population.

Reference List

(1) Dalziel TK. Thomas Kennedy Dalziel 1861-1924. Chronic interstitial enteritis. *Dis Colon Rectum* 1989;**32**(12):1076-8.

(2) Crohn BB, Ginzburg L, Oppenheimer GD. Landmark article Oct 15, 1932. Regional ileitis. A pathological and clinical entity. By Burril B. Crohn, Leon Ginzburg, and Gordon D. Oppenheimer. *JAMA* 1984;**251**(1):73-9.

(3) Penner A, Crohn BB. PERIANAL FISTULAE AS A COMPLICATION OF REGIONAL ILEITIS. *Ann Surg* 1938;**108**(5):867-73.

(4) Schofield PF. THE NATURAL HISTORY AND TREATMENT OF CROHN'S DISEASE. *Ann R Coll Surg Engl* 1965;**36**:258-79.

(5) D'Haens G, Baert F, Van Assche G *et al.* Early combined immunosuppression or conventional management in patients with newly diagnosed Crohn's disease: an open randomised trial. *Lancet* 2008;**371**(9613):660-7.

(6) Allchin WH. Ulcerative colitis- symposium and discussion based on 314 cases reported by the London hospitals. *R Soc Med* 1909;59-82.

(7) TRUELOVE SC, WITTS LJ. Cortisone in ulcerative colitis; final report on a therapeutic trial. *British Medical Journal (4947):1041-8,* 1955.

(8) Truelove SC. Medical management of ulcerative colitis. *Br Med J* 1968;**2**(5604):539-42.

(9) Roberts SE, Williams JG, Yeates D *et al.* Mortality in patients with and without colectomy admitted to hospital for ulcerative colitis and Crohn's disease: record linkage studies. *BMJ* 2007;**335**(7628):1033.

(10) Miller DS, Keighley AC, Langman MJ. Changing patterns in epidemiology of Crohn's disease. *Lancet* 1974;**2**(7882):691-3.

(11) Logan RF. Inflammatory bowel disease incidence: up, down or unchanged? *Gut* 1998;**42**(3):309-11.

(12) Economou M, Pappas G. New global map of Crohn's disease: Genetic, environmental, and socioeconomic correlations. *Inflamm Bowel Dis* 2008;**14**(5):709-20.

(13) Kyle J. Crohn's disease in the northeastern and northern Isles of Scotland: an epidemiological review. *Gastroenterology* 1992;**103**(2):392-9.

(14) Armitage E, Drummond H, Ghosh S *et al.* Incidence of juvenile-onset Crohn's disease in Scotland. *Lancet* 1999;**353**(9163):1496-7.

(15) Armitage EL, Aldhous MC, Anderson N *et al.* Incidence of juvenile-onset Crohn's disease in Scotland: association with northern latitude and affluence. *Gastroenterology* 2004;**127**(4):1051-7.

(16) Askling J, Grahnquist L, Ekbom A *et al.* Incidence of paediatric Crohn's disease in Stockholm, Sweden. *Lancet* 1999;**354**(9185):1179.

(17) Miller E, Waight P. Measles, measles vaccination, and Crohn's disease. Second immunisation has not affected incidence in England. *BMJ* 1998;**316**(7146):1745.

(18) Howarth GF, Robinson MH, Jenkins D *et al.* High prevalence of undetected ulcerative colitis: data from the Nottingham fecal occult blood screening trial. *Am J Gastroenterol* 2002;**97**(3):690-4.

(19) Armitage E, Drummond HE, Wilson DC *et al.* Increasing incidence of both juvenile-onset Crohn's disease and ulcerative colitis in Scotland. *Eur J Gastroenterol Hepatol* 2001;**13**(12):1439-47.

(20) Van Limbergen J, Russell RK, Drummond HE *et al.* Definition of Phenotypic Characteristics of Childhood-Onset Inflammatory Bowel Disease. *Gastroenterology* 2008.

(21) Cho JH. The genetics and immunopathogenesis of inflammatory bowel disease. *Nat Rev Immunol* 2008;**8**(6):458-66.

(22) Orholm M, Binder V, Sorensen TI *et al.* Concordance of inflammatory bowel disease among Danish twins. Results of a nationwide study. *Scand J Gastroenterol* 2000;**35**(10):1075-81.

(23) Thompson NP, Driscoll R, Pounder RE *et al.* Genetics versus environment in inflammatory bowel disease: results of a British twin study. *BMJ* 1996;**312**(7023):95-6.

(24) Tysk C, Lindberg E, Jarnerot G *et al.* Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. *Gut* 1988;**29**(7):990-6.

(25) Russell RK, Satsangi J. IBD: a family affair. *Best Pract Res Clin Gastroenterol* 2004;**18**(3):525-39.

(26) Tysk C, Lindberg E, Jarnerot G *et al.* Ulcerative colitis and Crohn's disease in an unselected population of monozygotic and dizygotic twins. A study of heritability and the influence of smoking. *Gut* 1988;**29**(7):990-6.

(27) Mayberry JF, Rhodes J, Newcombe RG. Familial prevalence of inflammatory bowel disease in relatives of patients with Crohn's disease. *Br Med J* 1980;**280**(6207):84.

(28) Probert CS, Jayanthi V, Hughes AO *et al.* Prevalence and family risk of ulcerative colitis and Crohn's disease: an epidemiological study among Europeans and south Asians in Leicestershire. *Gut* 1993;**34**(11):1547-51.

(29) Satsangi J, Rosenberg WMC, Jewell DP. The prevalence of inflammatory bowel disease in relatives of patients with Crohn's diseae. *European Journal of Gastroenterology and Hepatlogy* 1994;**6**:413-6.

(30) Weterman IT, Pena AS. Familial incidence of Crohn's disease in The Netherlands and a review of the literature. *Gastroenterology* 1984;**86**(3):449-52.

(31) Orholm M, Munkholm P, Langholz E *et al.* Familial occurrence of inflammatory bowel disease. *N Engl J Med* 1991;**324**(2):84-8.

(32) Orholm M, Iselius L, Sorensen TI *et al.* Investigation of inheritance of chronic inflammatory bowel diseases by complex segregation analysis. *BMJ* 1993;**306**(6869):20-4.

(33) Bonen DK, Cho JH. The genetics of inflammatory bowel disease. *Gastroenterology* 2003;**124**(2):521-36.

(34) Gaya DR, Russell RK, Nimmo ER *et al.* New genes in inflammatory bowel disease: lessons for complex diseases? *Lancet* 2006;**367**(9518):1271-84.

(35) Xavier RJ, Podolsky DK. Unravelling the pathogenesis of inflammatory bowel disease. *Nature* 2007;**448**(7152):427-34.

(36) Niess JH, Brand S, Gu X *et al.* CX3CR1-mediated dendritic cell access to the intestinal lumen and bacterial clearance. *Science* 2005;**307**(5707):254-8.

(37) Medzhitov R. Toll-like receptors and innate immunity. *Nat Rev Immunol* 2001;**1**(2):135-45.

(38) Meyers BC, Dickerman AW, Michelmore RW *et al.* Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. *Plant J* 1999;**20**(3):317-32.

(39) Hoffmann JA. The immune response of Drosophila. *Nature* 2003;**426**(6962):33-8.

(40) Akira S, Takeda K. Toll-like receptor signalling. *Nat Rev Immunol* 2004;**4**(7):499-511.

(41) Takeda K, Kaisho T, Akira S. Toll-like receptors. *Annu Rev Immunol* 2003;**21**:335-76.

(42) Hugot JP, Chamaillard M, Zouali H *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001;**411**(6837):599-603.

(43) Ogura Y, Bonen DK, Inohara N *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 2001;**411**(6837):603-6.

(44) Inohara N, Chamaillard M, McDonald C *et al.* NOD-LRR Proteins: Role in Host-Microbial Interactions and Inflammatory Disease. *Annu Rev Biochem* 2004.

(45) Ting JP, Williams KL. The CATERPILLER family: an ancient family of immune/apoptotic proteins. *Clin Immunol* 2005;**115**(1):33-7.

(46) Russell RK, Nimmo ER, Satsangi J. Molecular genetics of Crohn's disease. *Curr Opin Genet Dev* 2004;**14**(3):264-70.

(47) Girardin SE, Boneca IG, Viala J *et al.* Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. *J Biol Chem* 2003;**278**(11):8869-72.

(48) Inohara N, Ogura Y, Fontalba A *et al.* Host recognition of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease. *J Biol Chem* 2003;**278**(8):5509-12.

(49) Watanabe T, Kitani A, Murray PJ *et al.* NOD2 is a negative regulator of Toll-like receptor 2-mediated T helper type 1 responses. *Nat Immunol* 2004;**5**(8):800-8.

(50) Kobayashi KS, Chamaillard M, Ogura Y *et al.* Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. *Science* 2005;**307**(5710):731-4.

(51) Maeda S, Hsu LC, Liu H *et al.* Nod2 mutation in Crohn's disease potentiates NF-kappaB activity and IL-1beta processing. *Science* 2005;**307**(5710):734-8.

(52) Lees CW, Satsangi J. Autophagy and Crohn's Disease. *Inflammatory Bowl Disease Monitor* 2009;**Vol 9**(No 2).

(53) Sanjuan MA, Dillon CP, Tait SW *et al.* Toll-like receptor signalling in macrophages links the autophagy pathway to phagocytosis. *Nature* 2007;**450**(7173):1253-7.

(54) Cho JH, Weaver CT. The genetics of inflammatory bowel disease. *Gastroenterology* 2007;**133**(4):1327-39.

(55) LeibundGut-Landmann S, Gross O, Robinson MJ *et al.* Syk- and CARD9-dependent coupling of innate immunity to the induction of T helper cells that produce interleukin 17. *Nat Immunol* 2007;**8**(6):630-8.

(56) Napolitani G, Rinaldi A, Bertoni F *et al.* Selected Toll-like receptor agonist combinations synergistically trigger a T helper type 1-polarizing program in dendritic cells. *Nat Immunol* 2005;**6**(8):769-76.

(57) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;**447**(7145):661-78.

(58) Yen D, Cheung J, Scheerens H *et al.* IL-23 is essential for T cell-mediated colitis and promotes inflammation via IL-17 and IL-6. *J Clin Invest* 2006;**116**(5):1310-6.

(59) Sandborn WJ, Feagan BG, Fedorak RN *et al.* A randomized trial of Ustekinumab, a human interleukin-12/23 monoclonal antibody, in patients with moderate-to-severe Crohn's disease. *Gastroenterology* 2008;**135**(4):1130-41.

(60) Kim SC, Tonkonogy SL, Albright CA *et al.* Variable phenotypes of enterocolitis in interleukin 10-deficient mice monoassociated with two different commensal bacteria. *Gastroenterology* 2005;**128**(4):891-906.

(61) Barnich N, Darfeuille-Michaud A. Adherent-invasive Escherichia coli and Crohn's disease. *Curr Opin Gastroenterol* 2007;**23**(1):16-20.

(62) Sartor RB. Microbial influences in inflammatory bowel diseases. *Gastroenterology* 2008;**134**(2):577-94.

(63) Zeissig S, Burgel N, Gunzel D *et al.* Changes in expression and distribution of claudin 2, 5 and 8 lead to discontinuous tight junctions and barrier dysfunction in active Crohn's disease. *Gut* 2007;**56**(1):61-72.

(64) Panwala CM, Jones JC, Viney JL. A novel model of inflammatory bowel disease: mice deficient for the multiple drug resistance gene, mdr1a, spontaneously develop colitis. *J Immunol* 1998;**161**(10):5733-44.

(65) Ho GT, Soranzo N, Nimmo ER *et al.* ABCB1/MDR1 gene determines susceptibility and phenotype in ulcerative colitis: discrimination of critical variants using a gene-wide haplotype tagging approach. *Hum Mol Genet* 2006;**15**(5):797-805.

(66) Wehkamp J, Salzman NH, Porter E *et al.* Reduced Paneth cell alpha-defensins in ileal Crohn's disease. *Proc Natl Acad Sci U S A* 2005;**102**(50):18129-34.

(67) Wehkamp J, Harder J, Weichenthal M *et al.* NOD2 (CARD15) mutations in Crohn's disease are associated with diminished mucosal alpha-defensin expression. *Gut* 2004;**53**(11):1658-64.

(68) Simms LA, Doecke JD, Walsh MD *et al.* Reduced alpha-defensin expression is associated with inflammation and not NOD2 mutation status in ileal Crohn's disease. *Gut* 2008;**57**(7):903-10.

(69) A haplotype map of the human genome. *Nature* 2005;**437**(7063):1299-320.

(70) Duerr RH. Inflammatory Bowel Diseases. 1 ed ed. Curchill Livingstone, 2003.

(71) Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 1995;**11**(3):241-7.

(72) Cho JH, Nicolae DL, Gold LH *et al.* Identification of novel susceptibility loci for inflammatory bowel disease on chromosomes 1p, 3q, and 4q: evidence for epistasis between 1p and IBD1. *Proc Natl Acad Sci U S A* 1998;**95**(13):7502-7.

(73) Duerr RH, Barmada MM, Zhang L *et al.* Evidence for an inflammatory bowel disease locus on chromosome 3p26: linkage, transmission/disequilibrium and partitioning of linkage. *Hum Mol Genet* 2002;**11**(21):2599-606.

(74) Hampe J, Lynch NJ, Daniels S *et al.* Fine mapping of the chromosome 3p susceptibility locus in inflammatory bowel disease. *Gut* 2001;**48**(2):191-7.

(75) Ma Y, Ohmen JD, Li Z *et al.* A genome-wide search identifies potential new susceptibility loci for Crohn's disease. *Inflamm Bowel Dis* 1999;**5**(4):271-8.

(76) Rioux JD, Silverberg MS, Daly MJ *et al.* Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. *Am J Hum Genet* 2000;**66**(6):1863-70.

(77) Rioux JD, Daly MJ, Silverberg MS *et al.* Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat Genet* 2001;**29**(2):223-8.

(78) Hampe J, Schreiber S, Shaw SH *et al.* A genomewide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. *Am J Hum Genet* 1999;**64**(3):808-16.

(79) Satsangi J, Parkes M, Louis E *et al.* Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. *Nat Genet* 1996;**14**(2):199-202.

(80) Duerr RH, Barmada MM, Zhang L *et al.* High-density genome scan in Crohn disease shows confirmed linkage to chromosome 14q11-12. *Am J Hum Genet* 2000;**66**(6):1857-62.

(81) Hugot JP, Laurent-Puig P, Gower-Rousseau C *et al.* Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature* 1996;**379**(6568):821-3.

(82) Lesage S, Zouali H, Cezard JP *et al.* CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. *Am J Hum Genet* 2002;**70**(4):845-57.

(83) Inoue N, Tamura K, Kinouchi Y *et al.* Lack of common NOD2 variants in Japanese patients with Crohn's disease. *Gastroenterology* 2002;**123**(1):86-91.

(84) Leong RW, Armuzzi A, Ahmad T *et al.* NOD2/CARD15 gene polymorphisms and Crohn's disease in the Chinese population. *Aliment Pharmacol Ther* 2003;**17**(12):1465-70.

(85) Arnott ID, Nimmo ER, Drummond HE *et al.* NOD2/CARD15, TLR4 and CD14 mutations in Scottish and Irish Crohn's disease patients: evidence for genetic heterogeneity within Europe? *Genes Immun* 2004;**5**(5):417-25.

(86) Torkvist L, Noble CL, Lordal M *et al.* Contribution of CARD15 variants in determining susceptibility to Crohn's disease in Sweden. *Scand J Gastroenterol* 2006;**41**(6):700-5.

(87) Ahmad T, Armuzzi A, Bunce M *et al.* The molecular classification of the clinical manifestations of Crohn's disease. *Gastroenterology* 2002;**122**(4):854-66.

(88) Economou M, Trikalinos TA, Loizou KT *et al.* Differential effects of NOD2 variants on Crohn's disease risk and phenotype in diverse populations: a metaanalysis. *Am J Gastroenterol* 2004;**99**(12):2393-404.

(89) Daly MJ, Rioux JD, Schaffner SF *et al.* High-resolution haplotype structure in the human genome. *Nat Genet* 2001;**29**(2):229-32.

(90) Giallourakis C, Stoll M, Miller K *et al.* IBD5 is a general risk factor for inflammatory bowel disease: replication of association with Crohn disease and identification of a novel association with ulcerative colitis. *Am J Hum Genet* 2003;**73**(1):205-11.

(91) Mirza MM, Fisher SA, King K *et al.* Genetic evidence for interaction of the 5q31 cytokine locus and the CARD15 gene in Crohn disease. *Am J Hum Genet* 2003;**72**(4):1018-22.

(92) Negoro K, McGovern DP, Kinouchi Y *et al.* Analysis of the IBD5 locus and potential gene-gene interactions in Crohn's disease. *Gut* 2003;**52**(4):541-6.

(93) Armuzzi A, Ahmad T, Ling KL *et al.* Genotype-phenotype analysis of the Crohn's disease susceptibility haplotype on chromosome 5q31. *Gut* 2003;**52**(8):1133-9.

(94) Tosa M, Negoro K, Kinouchi Y *et al.* Lack of association between IBD5 and Crohn's disease in Japanese patients demonstrates population-specific differences in inflammatory bowel disease. *Scand J Gastroenterol* 2006;**41**(1):48-53.

(95) Peltekova VD, Wintle RF, Rubin LA *et al.* Functional variants of OCTN cation transporter genes are associated with Crohn disease. *Nat Genet* 2004;**36**(5):471-5.

(96) Noble CL, Nimmo ER, Drummond H *et al.* The contribution of OCTN1/2 variants within the IBD5 locus to disease susceptibility and severity in Crohn's disease. *Gastroenterology* 2005;**129**(6):1854-64.

(97) Torok HP, Glas J, Tonenchi L *et al.* Polymorphisms in the DLG5 and OCTN cation transporter genes in Crohn's disease. *Gut* 2005;**54**(10):1421-7.

(98) Russell RK, Drummond HE, Nimmo ER *et al.* Analysis of the influence of OCTN1/2 variants within the IBD5 locus on disease susceptibility and growth indices in early onset inflammatory bowel disease. *Gut* 2006;**55**(8):1114-23.

(99) Waller S, Tremelling M, Bredin F *et al.* Evidence for association of OCTN genes and IBD5 with ulcerative colitis. *Gut* 2006;**55**(6):809-14.

(100) Fisher SA, Hampe J, Onnie CM *et al.* Direct or indirect association in a complex disease: the role of SLC22A4 and SLC22A5 functional variants in Crohn disease. *Hum Mutat* 2006;**27**(8):778-85.

(101) Tamai I, Yabuuchi H, Nezu J *et al.* Cloning and characterization of a novel human pH-dependent organic cation transporter, OCTN1. *FEBS Lett* 1997;**419**(1):107-11.

(102) Tokuhiro S, Yamada R, Chang X *et al.* An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. *Nat Genet* 2003;**35**(4):341-8.

(103) Tamai I, Ohashi R, Nezu J *et al.* Molecular and functional identification of sodium ion-dependent, high affinity human carnitine transporter OCTN2. *J Biol Chem* 1998;**273**(32):20378-82.

(104) Lohoff M, Mak TW. Roles of interferon-regulatory factors in T-helper-cell differentiation. *Nat Rev Immunol* 2005;**5**(2):125-35.

(105) Clavell M, Correa-Gracian H, Liu Z *et al.* Detection of interferon regulatory factor-1 in lamina propria mononuclear cells in Crohn's disease. *J Pediatr Gastroenterol Nutr* 2000;**30**(1):43-7.

(106) Reinhard C, Rioux JD. Role of the IBD5 susceptibility locus in the inflammatory bowel diseases. *Inflamm Bowel Dis* 2006;**12**(3):227-38.

(107) Stoll M, Corneliussen B, Costello CM *et al.* Genetic variation in DLG5 is associated with inflammatory bowel disease. *Nat Genet* 2004;**36**(5):476-80.

(108) van den Berk LC, van Ham MA, te Lindert MM *et al.* The interaction of PTP-BL PDZ domains with RIL: an enigmatic role for the RIL LIM domain. *Mol Biol Rep* 2004;**31**(4):203-15.

(109) Dieckgraefe BK, Korzenik JR. Treatment of active Crohn's disease with recombinant human granulocyte-macrophage colony-stimulating factor. *Lancet* 2002;**360**(9344):1478-80.

(110) Korzenik JR, Dieckgraefe BK, Valentine JF *et al.* Sargramostim for active Crohn's disease. *N Engl J Med* 2005;**352**(21):2193-201.

(111) Valentine JF, Fedorak RN, Feagan B *et al.* Steroid-sparing properties of sargramostim in patients with corticosteroid-dependent Crohn's disease: a

randomised, double-blind, placebo-controlled, phase 2 study. *Gut* 2009;**58**(10):1354-62.

(112) Newman B, Gu X, Wintle R *et al.* A risk haplotype in the solute carrier family 22A4/22A5 gene cluster influences phenotypic expression of Crohn's disease. *Gastroenterology* 2005;**128**(2):260-9.

(113) Vermeire S, Pierik M, Hlavaty T *et al.* Association of organic cation transporter risk haplotype with perianal penetrating Crohn's disease but not with susceptibility to IBD. *Gastroenterology* 2005;**129**(6):1845-53.

(114) Silverberg MS, Satsangi J, Ahmad T *et al.* Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: Report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. *Can J Gastroenterol* 2005;**19 Suppl A**:5-36.

(115) Gasche C, Scholmerich J, Brynskov J *et al.* A simple classification of Crohn's disease: report of the Working Party for the World Congresses of Gastroenterology, Vienna 1998. *Inflamm Bowel Dis* 2000;**6**(1):8-15.

(116) Palmieri O, Latiano A, Valvano R *et al.* Variants of OCTN1-2 cation transporter genes are associated with both Crohn's disease and ulcerative colitis. *Aliment Pharmacol Ther* 2006;**23**(4):497-506.

(117) Urcelay E, Mendoza JL, Martinez A *et al.* IBD5 polymorphisms in inflammatory bowel disease: association with response to infliximab. *World J Gastroenterol* 2005;**11**(8):1187-92.

(118) Palmieri O, Latiano A, Valvano R *et al.* Variants of OCTN1-2 cation transporter genes are associated with both Crohn's disease and ulcerative colitis. *Aliment Pharmacol Ther* 2006;**23**(4):497-506.

(119) McGovern DP, Van Heel DA, Negoro K *et al.* Further evidence of IBD5/CARD15 (NOD2) epistasis in the susceptibility to ulcerative colitis. *Am J Hum Genet* 2003;**73**(6):1465-6.

(120) Daly MJ, Pearce AV, Farwell L *et al.* Association of DLG5 R30Q variant with inflammatory bowel disease. *Eur J Hum Genet* 2005;**13**(7):835-9.

(121) Tenesa A, Noble C, Satsangi J *et al.* Association of DLG5 and inflammatory bowel disease across populations. *Eur J Hum Genet* 2006;**14**(3):259-60.

(122) Noble CL, Nimmo ER, Drummond H *et al.* DLG5 variants do not influence susceptibility to inflammatory bowel disease in the Scottish population. *Gut* 2005.

(123) Newman WG, Gu X, Wintle RF *et al.* DLG5 variants contribute to Crohn disease risk in a Canadian population. *Hum Mutat* 2006;**27**(4):353-8.

(124) Medici V, Mascheretti S, Croucher PJ *et al.* Extreme heterogeneity in CARD15 and DLG5 Crohn disease-associated polymorphisms between German and Norwegian populations. *Eur J Hum Genet* 2006;**14**(4):459-68.

(125) Buning C, Geerdts L, Fiedler T *et al.* DLG5 variants in inflammatory bowel disease. *Am J Gastroenterol* 2006;**101**(4):786-92.

(126) Lakatos PL, Fischer S, Claes K *et al.* DLG5 R30Q is not associated with IBD in Hungarian IBD patients but predicts clinical response to steroids in Crohn's disease. *Inflamm Bowel Dis* 2006;**12**(5):362-8.

(127) Tremelling M, Waller S, Bredin F *et al.* Genetic variants in TNF-alpha but not DLG5 are associated with inflammatory bowel disease in a large United Kingdom cohort. *Inflamm Bowel Dis* 2006;**12**(3):178-84.

(128) Pearce AV, Fisher SA, Prescott NJ *et al.* Investigation of association of the DLG5 gene with phenotypes of inflammatory bowel disease in the British population. *Int J Colorectal Dis* 2007;**22**(4):419-24.

(129) Lakatos PL, Fischer S, Claes K *et al.* DLG5 R30Q is not associated with IBD in Hungarian IBD patients but predicts clinical response to steroids in Crohn's disease. *Inflamm Bowel Dis* 2006;**12**(5):362-8.

(130) Yamazaki K, Takazoe M, Tanaka T *et al.* Association analysis of SLC22A4, SLC22A5 and DLG5 in Japanese patients with Crohn disease. *J Hum Genet* 2004;**49**(12):664-8.

(131) Gazouli M, Mantzaris G, Archimandritis AJ *et al.* Single nucleotide polymorphisms of OCTN1, OCTN2, and DLG5 genes in Greek patients with Crohn's disease. *World J Gastroenterol* 2005;**11**(47):7525-30.

(132) Tosa M, Negoro K, Kinouchi Y *et al.* Lack of association between IBD5 and Crohn's disease in Japanese patients demonstrates population-specific differences in inflammatory bowel disease. *Scand J Gastroenterol* 2006;**41**(1):48-53.

(133) Lees CW, Nimmo ER, Russell RK *et al.* Analysis of CCL20 variants in ibd provides further evidence for genetic heterogeneity in disease susceptibility. *Gut* 2006;**55**:A1.

(134) Russell RK, Drummond HE, Nimmo ER *et al.* The contribution of the DLG5 113A variant in early-onset inflammatory bowel disease. *J Pediatr* 2007;**150**(3):268-73.

(135) Friedrichs F, Brescianini S, Annese V *et al.* Evidence of transmission ratio distortion of DLG5 R30Q variant in general and implication of an association with Crohn disease in men. *Hum Genet* 2006;**119**(3):305-11.

(136) Browning BL, Barclay ML, Bingham SA *et al.* Gender-stratified analysis of DLG5 R30Q in 4707 Crohn's disease patients and 4973 controls from 12 Caucasian cohorts. *J Med Genet* 2007.

(137) Moore JH. A global view of epistasis. *Nat Genet* 2005;**37**(1):13-4.

(138) Gonzalez-Mariscal L, Betanzos A, Avila-Flores A. MAGUK proteins: structure and role in the tight junction. *Semin Cell Dev Biol* 2000;**11**(4):315-24.

(139) Dimitratos SD, Woods DF, Stathakis DG *et al.* Signaling pathways are focused at specialized regions of the plasma membrane by scaffolding proteins of the MAGUK family. *Bioessays* 1999;**21**(11):912-21.

(140) Stehle T, Schulz GE. Refined structure of the complex between guanylate kinase and its substrate GMP at 2.0 A resolution. *J Mol Biol* 1992;**224**(4):1127-41.

(141) Shah G, Brugada R, Gonzalez O *et al.* The cloning, genomic organization and tissue expression profile of the human DLG5 gene. *BMC Genomics* 2002;**3**(1):6.

(142) Friedrichs F, Henckaerts L, Vermeire S *et al.* The Crohn's disease susceptibility gene DLG5 as a member of the CARD interaction network. *J Mol Med* 2008.

(143) Pe'er I, de Bakker PI, Maller J *et al.* Evaluating and improving power in whole-genome association studies using fixed marker sets. *Nat Genet* 2006;**38**(6):663-7.

(144) Yamazaki K, McGovern D, Ragoussis J *et al.* Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn's disease. *Hum Mol Genet* 2005;**14**(22):3499-506.

(145) Duerr RH, Taylor KD, Brant SR *et al.* A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* 2006;**314**(5804):1461-3.

(146) Tremelling M, Cummings F, Fisher SA *et al.* IL23R variation determines susceptibility but not disease phenotype in inflammatory bowel disease. *Gastroenterology* 2007;**132**(5):1657-64.

(147) Libioulle C, Louis E, Hansoul S *et al.* Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLoS Genet* 2007;**3**(4):e58.

(148) Raelson JV, Little RD, Ruether A *et al.* Genome-wide association study for Crohn's disease in the Quebec Founder Population identifies multiple validated disease loci. *Proc Natl Acad Sci U S A* 2007;**104**(37):14747-52.

(149) Parkes M, Barrett JC, Prescott NJ *et al.* Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat Genet* 2007;**39**(7):830-2.

(150) Rioux JD, Xavier RJ, Taylor KD *et al.* Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet* 2007;**39**(5):596-604.

(151) Barrett JC, Hansoul S, Nicolae DL *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* 2008.

(152) Fisher SA, Tremelling M, Anderson CA *et al.* Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease. *Nat Genet* 2008;**40**(6):710-2.

(153) Silverberg MS, Cho JH, Rioux JD *et al.* Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. *Nat Genet* 2009;**41**(2):216-20.

(154) Franke A, Balschun T, Karlsen TH *et al.* Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. *Nat Genet* 2008;**40**(11):1319-23.

(155) Stoughton RB. Applications of DNA microarrays in biology. *Annu Rev Biochem* 2005;**74**:53-82.

(156) Schena M, Shalon D, Davis RW *et al.* Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995;**270**(5235):467-70.

(157) Marshall E. Getting the noise out of gene arrays. *Science* 2004;**306**(5696):630-1.

(158) Warner EE, Dieckgraefe BK. Application of genome-wide gene expression profiling by high-density DNA arrays to the treatment and study of inflammatory bowel disease. *Inflamm Bowel Dis* 2002;**8**(2):140-57.

(159) Tilstone C. DNA microarrays: vital statistics. *Nature* 2003;**424**(6949):610-2.

(160) Phimister B. Going global. *Nature Genetics* 1999;**21**:1.

(161) Affymetrix home page. 2006.

(162) Microarray figure. *http://www eng usf edu/~das/CancerBiology/images/microarray bmp* 2006.

(163) Wu TD. Analysing gene expression data from DNA microarrays to identify candidate genes. *J Pathol* 2001;**195**(1):53-65.

(164) Agilent Whole Human Genome Oligo Microarray Kit. *http://www chem agilent com/temp/rad429DA/00048373 pdf* 2006.

(165) Heller RA, Schena M, Chai A *et al.* Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc Natl Acad Sci U S A* 1997;**94**(6):2150-5

(166) Dieckgraefe BK, Stenson WF, Korzenik JR *et al.* Analysis of mucosal gene expression in inflammatory bowel disease by parallel oligonucleotide arrays. *Physiol Genomics* 2000;**4**(1):1-11.

(167) Lawrance IC, Fiocchi C, Chakravarti S. Ulcerative colitis and Crohn's disease: distinctive gene expression profiles and novel susceptibility candidate genes. *Hum Mol Genet* 2001;**10**(5):445-56.

(168) Langmann T, Moehle C, Mauerer R *et al.* Loss of detoxification in inflammatory bowel disease: dysregulation of pregnane X receptor target genes. *Gastroenterology* 2004;**127**(1):26-40.

(169) Ioannidis JP. Microarrays and molecular research: noise discovery? *Lancet* 2005;**365**(9458):454-5.

(170) Eisenstein M. Microarrays: quality control. *Nature* 2006;**442**(7106):1067-70.

(171) Agilent Whole Human Genome Oligo Microarray Kit. 2006.

(172) Hughes TR, Mao M, Jones AR *et al.* Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* 2001;**19**(4):342

(173) Kamme F, Zhu J, Luo L *et al.* Single-cell laser-capture microdissection and RNA amplification. *Methods Mol Med* 2004;**99**:215-23.

(174) Brazma A, Hingamp P, Quackenbush J *et al.* Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 2001;**29**(4):365-71.

(175) Durbin BP, Rocke DM. Variance-stabilizing transformations for two-color microarrays. *Bioinformatics* 2004;**20**(5):660-7.

(176) Hughes TR, Marton MJ, Jones AR *et al.* Functional discovery via a compendium of expression profiles. *Cell* 2000;**102**(1):109-26.

(177) Ji H, Davis RW. Data quality in genomics and microarrays. *Nat Biotechnol* 2006;**24**(9):1112-3.

(178) Baker SC, Bauer SR, Beyer RP *et al.* The External RNA Controls Consortium: a progress report. *Nat Methods* 2005;**2**(10):731-4.

(179) Shi L, Reid LH, Jones WD *et al.* The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 2006;**24**(9):1151-61.

(180) Patterson TA, Lobenhofer EK, Fulmer-Smentek SB *et al.* Performance comparison of one-color and two-color platforms within the Microarray Quality Control (MAQC) project. *Nat Biotechnol* 2006;**24**(9):1140-50.

(181) Canales RD, Luo Y, Willey JC *et al.* Evaluation of DNA microarray results with quantitative gene expression platforms. *Nat Biotechnol* 2006;**24**(9):1115-22.

(182) Hoheisel JD. Microarray technology: beyond transcript profiling and genotype analysis. *Nat Rev Genet* 2006;**7**(3):200-10.

(183) Tilstone C. DNA microarrays: vital statistics. *Nature* 2003;**424**(6949):610-2.

(184) Benjamini Y, Drai D, Elmer G *et al.* Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* 2001;**125**(1-2):279-84.

(185) Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 2003;**100**(16):9440-5.

(186) Eisenstein M. Quality control. *Nature* 2006;**442**(7106):1067-70.

(187) Spasic I, Ananiadou S, McNaught J *et al.* Text mining and ontologies in biomedicine: making sense of raw text. *Brief Bioinform* 2005;**6**(3):239-51.

(188) Hoheisel JD. Microarray technology: beyond transcript profiling and genotype analysis. *Nat Rev Genet* 2006;**7**(3):200-10.

(189) Alizadeh AA, Eisen MB, Davis RE *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000;**403**(6769):503-11.

(190) van de Vijver MJ, He YD, Van't Veer LJ *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;**347**(25):1999-2009.

(191) Ntzani EE, Ioannidis JP. Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. *Lancet* 2003;**362**(9394):1439-44.

(192) Diosdado B, Wapenaar MC, Franke L *et al.* A microarray screen for novel candidate genes in coeliac disease pathogenesis. *Gut* 2004;**53**(7):944-51.

(193) Juuti-Uusitalo K, Maki M, Kaukinen K *et al.* cDNA microarray analysis of gene expression in coeliac disease jejunal biopsy samples. *J Autoimmun* 2004;**22**(3):249-65.

(194) Heller RA, Schena M, Chai A *et al.* Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc Natl Acad Sci U S A* 1997;**94**(6):2150-5.

(195) Okahara S, Arimura Y, Yabana T *et al.* Inflammatory gene signature in ulcerative colitis with cDNA macroarray analysis. *Aliment Pharmacol Ther* 2005;**21**(9):1091-7.

(196) Costello CM, Mah N, Hasler R *et al.* Dissection of the inflammatory bowel disease transcriptome using genome-wide cDNA microarrays. *PLoS Med* 2005;**2**(8):e199.

(197) Puleston J, Cooper M, Murch S *et al.* A distinct subset of chemokines dominates the mucosal chemokine response in inflammatory bowel disease. *Aliment Pharmacol Ther* 2005;**21**(2):109-20.

(198) Burczynski ME, Peterson RL, Twine NC *et al.* Molecular classification of Crohn's disease and ulcerative colitis patients using transcriptional profiles in peripheral blood mononuclear cells. *J Mol Diagn* 2006;**8**(1):51-61.

(199) Li J, Moran T, Swanson E *et al.* Regulation of IL-8 and IL-1beta expression in Crohn's disease associated NOD2/CARD15 mutations. *Hum Mol Genet* 2004;**13**(16):1715-25.

(200) Sauer S, Lange BM, Gobom J *et al.* Miniaturization in functional genomics and proteomics. *Nat Rev Genet* 2005;**6**(6):465-76.

(201) Din S, Lennon AM, Hogarth C *et al.* Proteomic Profiling Identifies Corticosteroid Resistant Patients In Severe Ulcerative Colitis. *Gastroenterology* 2005;**128**(4):A237.

(202) Alberts B, Johnson A, Lewis J *et al.* Molecular Biology of the Cell. 4th Edition ed. Garland Publishing, 2002.

(203) Lodish H, Berk A, Zipursky S L *et al.* Molecular Cell Biology. 4 th ed ed. W. H. Freedman and Co, 2000.

(204) Millar JK, Wilson-Annan JC, Anderson S *et al.* Disruption of two novel genes by a translocation co-segregating with schizophrenia. *Hum Mol Genet* 2000;**9**(9):1415-23.

(205) Blackwood DH, Fordyce A, Walker MT *et al.* Schizophrenia and affective disorders--cosegregation with a translocation at chromosome 1q42 that directly disrupts brain-expressed genes: clinical and P300 findings in a family. *Am J Hum Genet* 2001;**69**(2):428-33.

(206) Ekelund J, Hovatta I, Parker A *et al.* Chromosome 1 loci in Finnish schizophrenia families. *Hum Mol Genet* 2001;**10**(15):1611-7.

(207) Millar JK, Christie S, Porteous DJ. Yeast two-hybrid screens implicate DISC1 in brain development and function. *Biochem Biophys Res Commun* 2003;**311**(4):1019-25.

(208) Morris JA, Kandpal G, Ma L *et al.* DISC1 (Disrupted-In-Schizophrenia 1) is a centrosome-associated protein that interacts with MAP1A, MIPT3, ATF4/5 and NUDEL: regulation and loss of interaction with mutation. *Hum Mol Genet* 2003;**12**(13):1591-608.

(209) Ozeki Y, Tomoda T, Kleiderlein J *et al.* Disrupted-in-Schizophrenia-1 (DISC-1): mutant truncation prevents binding to NudE-like (NUDEL) and inhibits neurite outgrowth. *Proc Natl Acad Sci U S A* 2003;**100**(1):289-94.

(210) Barnich N, Aguirre JE, Reinecker HC *et al.* Membrane recruitment of NOD2 in intestinal epithelial cells is essential for nuclear factor-{kappa}B activation in muramyl dipeptide recognition. *J Cell Biol* 2005;**170**(1):21-6.

(211) Lennard-Jones JE. Classification of inflammatory bowel disease. *Scand J Gastroenterol Suppl* 1989;**170**:2-6.

(212) Walmsley RS, Ayres RC, Pounder RE *et al.* A simple clinical colitis activity index. *Gut* 1998;**43**(1):29-32.

(213) Harvey RF, Bradshaw JM. A simple index of Crohn's-disease activity. *Lancet* 1980;**1**(8167):514.

(214) Jubb AM, Pham TQ, Frantz GD *et al.* Quantitative in situ hybridization of tissue microarrays. *Methods Mol Biol* 2006;**326**:255-64.

(215) Schlesselman JJ. Case Control Studies: Design, Conduct, Analysis. Stolley PD. 1982. Oxford Press.

(216) Maloy KJ, Kullberg MC. IL-23 and Th17 cytokines in intestinal homeostasis. *Mucosal Immunol* 2008;**1**(5):339-49.

(217) Abbas AR, Baldwin D, Ma Y *et al.* Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. *Genes Immun* 2005;**6**(4):319-31.

(218) Dwinell MB, Lugering N, Eckmann L *et al.* Regulated production of interferon-inducible T-cell chemoattractants by human intestinal epithelial cells. *Gastroenterology* 2001;**120**(1):49-59.

(219) Lee JW, Wang P, Kattah MG *et al.* Differential regulation of chemokines by IL-17 in colonic epithelial cells. *J Immunol* 2008;**181**(9):6536-45.

(220) Yang SK, Eckmann L, Panja A *et al.* Differential and regulated expression of C-X-C, C-C, and C-chemokines by human colon epithelial cells. *Gastroenterology* 1997;**113**(4):1214-23.

(221) Gutfeld O, Prus D, Ackerman Z *et al.* Expression of serum amyloid A, in normal, dysplastic, and neoplastic human colonic mucosa: implication for a role in colonic tumorigenesis. *J Histochem Cytochem* 2006;**54**(1):63-73.

(222) Ray A, Shakya A, Kumar D *et al.* Inflammation-responsive transcription factor SAF-1 activity is linked to the development of amyloid A amyloidosis. *J Immunol* 2006;**177**(4):2601-9.

(223) Madsen L, Schulze A, Seeger M *et al.* Ubiquitin domain proteins in disease. *BMC Biochem* 2007;**8 Suppl 1**:S1.

(224) Fan W, Cai W, Parimoo S *et al.* Identification of seven new human MHC class I region genes around the HLA-F locus. *Immunogenetics* 1996;**44**(2):97-103.

(225) Lee CG, Ren J, Cheong IS *et al.* Expression of the FAT10 gene is highly upregulated in hepatocellular carcinoma and other gastrointestinal and gynecological cancers. *Oncogene* 2003;**22**(17):2592-603.

(226) Ren J, Kan A, Leong SH *et al.* FAT10 plays a role in the regulation of chromosomal stability. *J Biol Chem* 2006;**281**(16):11413-21.

(227) Zhang DW, Jeang KT, Lee CG. p53 negatively regulates the expression of FAT10, a gene upregulated in various cancers. *Oncogene* 2006;**25**(16):2318-27.

(228) Lukasiak S, Schiller C, Oehlschlaeger P *et al.* Proinflammatory cytokines cause FAT10 upregulation in cancers of liver and colon. *Oncogene* 2008;**27**(46):6068-74.

(229) Simms LA, Doecke JD, Walsh MD *et al.* Reduced alpha-defensin expression is associated with inflammation and not NOD2 mutation status in ileal Crohn's disease. *Gut* 2008;**57**(7):903-10.

(230) Noble CL, Abbas AR, Cornelius J *et al.* Regional Variation in Gene Expression in the Healthy Colon is Dysregulated in Ulcerative Colitis. *Gut* 2008.

(231) Al Shami A, Spolski R, Kelly J *et al.* A role for TSLP in the development of inflammation in an asthma model. *J Exp Med* 2005;**202**(6):829-39.

(232) Zaph C, Troy AE, Taylor BC *et al.* Epithelial-cell-intrinsic IKK-beta expression regulates intestinal immune homeostasis. *Nature* 2007;**446**(7135):552-6.

(233) Sanjuan MA, Dillon CP, Tait SW *et al.* Toll-like receptor signalling in macrophages links the autophagy pathway to phagocytosis. *Nature* 2007;**450**(7173):1253-7.

(234) Pabst O, Forster R, Lipp M *et al.* NKX2.3 is required for MAdCAM-1 expression and homing of lymphocytes in spleen and mucosa-associated lymphoid tissue. *EMBO J* 2000;**19**(9):2015-23.

(235) Leung Y, Panaccione R. Anti-adhesion molecule strategies for Crohn disease. *BioDrugs* 2008;**22**(4):259-64.

(236) Puleston J, Cooper M, Murch S *et al.* A distinct subset of chemokines dominates the mucosal chemokine response in inflammatory bowel disease. *Aliment Pharmacol Ther* 2005;**21**(2):109-20.

(237) Banks C, Bateman A, Payne R *et al.* Chemokine expression in IBD. Mucosal chemokine expression is unselectively increased in both ulcerative colitis and Crohn's disease. *J Pathol* 2003;**199**(1):28-35.

(238) Kwon JH, Keates S, Bassani L *et al.* Colonic epithelial cells are a major site of macrophage inflammatory protein 3alpha (MIP-3alpha) production in normal colon and inflammatory bowel disease. *Gut* 2002;**51**(6):818-26.

(239) Franke A, Balschun T, Karlsen TH *et al.* Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. *Nat Genet* 2008;**40**(11):1319-23.

(240) De Santa BP, Roberts DJ. Tail gut endoderm and gut/genitourinary/tail development: a new tissue-specific role for Hoxa13. *Development* 2002;**129**(3):551-61.

(241) Jung C, Kim RS, Zhang H *et al.* HOXB13 is downregulated in colorectal cancer to confer TCF4-mediated transactivation. *Br J Cancer* 2005;**92**(12):2233-9.

(242) Lees C, Howie S, Sartor RB *et al.* The hedgehog signalling pathway in the gastrointestinal tract: implications for development, homeostasis, and disease. *Gastroenterology* 2005;**129**(5):1696-710.

(243) Parkes M, Barmada MM, Satsangi J *et al.* The IBD2 locus shows linkage heterogeneity between ulcerative colitis and Crohn disease. *Am J Hum Genet* 2000;**67**(6):1605-10.

(244) Lees CW, Zacharias WJ, Tremelling M *et al.* Analysis of germline GLI1 variation implicates hedgehog signalling in the regulation of intestinal inflammatory pathways. *PLoS Med* 2008;**5**(12):e239.

(245) Wu F, Dassopoulos T, Cope L *et al.* Genome-wide gene expression differences in Crohn's disease and ulcerative colitis from endoscopic pinch biopsies: insights into distinctive pathogenesis. *Inflamm Bowel Dis* 2007;**13**(7):807-21.

(246) Varnat F, Heggeler BB, Grisel P *et al.* PPARbeta/delta regulates paneth cell differentiation via controlling the hedgehog signaling pathway. *Gastroenterology* 2006;**131**(2):538-53.

(247) Ogawa H, Fukushima K, Naito H *et al.* Increased expression of HIP/PAP and regenerating gene III in human inflammatory bowel disease and a murine bacterial reconstitution model. *Inflamm Bowel Dis* 2003;**9**(3):162-70.

(248) Ho GT, Soranzo N, Nimmo ER *et al.* ABCB1/MDR1 gene determines susceptibility and phenotype in ulcerative colitis: discrimination of critical variants using a gene-wide haplotype tagging approach. *Hum Mol Genet* 2006;**15**(5):797-805.

(249) Ho GT, Soranzo N, Tate SK *et al.* Lack of association of the pregnane X receptor (PXR/NR1I2) gene with inflammatory bowel disease: parallel allelic association study and gene wide haplotype analysis. *Gut* 2006;**55**(11):1676-7.

(250) Staege H, Brauchlin A, Schoedon G *et al.* Two novel genes FIND and LIND differentially expressed in deactivated and Listeria-infected human macrophages. *Immunogenetics* 2001;**53**(2):105-13.

(251) Capelli C, Redhead N, Abernethy JK *et al.* A Y chromosome census of the British Isles. *Curr Biol* 2003;**13**(11):979-84.

(252) Torkvist L, Noble CL, Lordal M *et al.* Contribution of the IBD5 locus to Crohn's disease in the Swedish population. *Scand J Gastroenterol* 2007;**42**(2):200-6.

(253) Silverberg MS, Duerr RH, Brant SR *et al.* Refined genomic localization and ethnic differences observed for the IBD5 association with Crohn's disease. *Eur J Hum Genet* 2007;**15**(3):328-35.

(254) Urban TJ, Giacomini KM, Risch N. Haplotype structure and ethnic-specific allele frequencies at the OCTN locus: implications for the genetics of Crohn's disease. *Inflamm Bowel Dis* 2005;**11**(1):78-9.

(255) Lahjouji K, Mitchell GA, Qureshi IA. Carnitine transport by organic cation transporters and systemic carnitine deficiency. *Mol Genet Metab* 2001;**73**(4):287-97.

(256) Clavell M, Correa-Gracian H, Liu Z *et al.* Detection of interferon regulatory factor-1 in lamina propria mononuclear cells in Crohn's disease. *J Pediatr Gastroenterol Nutr* 2000;**30**(1):43-7.

(257) Lohoff M, Mak TW. Roles of interferon-regulatory factors in T-helper-cell differentiation. *Nat Rev Immunol* 2005;**5**(2):125-35.

(258) Sachar DB. Genomics and phenomics in Crohn's disease. *Gastroenterology* 2002;**122**(4):1161-2.

(259) Brescianini S, Trinh T, Stoll M *et al.* IBD5 is associated with an extensive complicated Crohn's disease feature: implications from genotype-phenotype analysis. *Gut* 2007;**56**(1):149-50.

(260) Bairead E, Harmon DL, Curtis AM *et al.* Association of NOD2 with Crohn's disease in a homogenous Irish population. *Eur J Hum Genet* 2003;**11**(3):237-44.

(261) Helio T, Halme L, Lappalainen M *et al.* CARD15/NOD2 gene variants are associated with familially occurring and complicated forms of Crohn's disease. *Gut* 2003;**52**(4):558-62.

(262) Sugimura M, Kinouchi Y, Takahashi S *et al.* CARD15/NOD2 mutational analysis in Japanese patients with Crohn's disease. *Clin Genet* 2003;**63**(2):160-2.

(263) Marsh S, McLeod HL. Crohn's disease: ethnic variation in CARD15 genotypes. *Gut* 2003;**52**(5):770.

(264) Franke A, Balschun T, Karlsen TH *et al.* Replication of signals from recent studies of Crohn's disease identifies previously unknown disease loci for ulcerative colitis. *Nat Genet* 2008;**40**(6):713-5.

(265) Arnott ID, Ho GT, Nimmo ER *et al.* Toll-like receptor 4 gene in IBD: further evidence for genetic heterogeneity in Europe * Author's reply. *Gut* 2005;**54**(2):308-9.

(266) Barton GM, Medzhitov R. Toll-like receptor signaling pathways. *Science* 2003;**300**(5625):1524-5.

(267) Franchimont D, Vermeire S, El Housni H *et al.* Deficient host-bacteria interactions in inflammatory bowel disease? The toll-like receptor (TLR)-4 Asp299gly polymorphism is associated with Crohn's disease and ulcerative colitis. *Gut* 2004;**53**(7):987-92.

(268) Torok HP, Glas J, Tonenchi L *et al.* Polymorphisms of the lipopolysaccharide-signaling complex in inflammatory bowel disease: association of a mutation in the Toll-like receptor 4 gene with ulcerative colitis. *Clin Immunol* 2004;**112**(1):85-91.

(269) Russell RK, Drummond HE, Nimmo ER *et al.* Socio-economic status critically influences penetrance of the DLG5 113A variant in early-onset IBD: A novel gene-environmental interaction. *Gastroenterology* 2006;**130**(4):A53.

(270) Brewerton DA, Caffrey M, Nicholls A *et al.* HL-A 27 and arthropathies associated with ulcerative colitis and psoriasis. *Lancet* 1974;**1**(7864):956-8.

(271) Orchard TR, Thiyagaraja S, Welsh KI *et al.* Clinical phenotype is related to HLA genotype in the peripheral arthropathies of inflammatory bowel disease. *Gastroenterology* 2000;**118**(2):274-8.

(272) Spurkland A, Saarinen S, Boberg KM *et al.* HLA class II haplotypes in primary sclerosing cholangitis patients from five European populations. *Tissue Antigens* 1999;**53**(5):459-69.

(273) Satsangi J, Chapman RW, Haldar N *et al.* A functional polymorphism of the stromelysin gene (MMP-3) influences susceptibility to primary sclerosing cholangitis. *Gastroenterology* 2001;**121**(1):124-30.

(274) Arijs I, Li K, Toedter G *et al.* Mucosal gene signatures to predict response to infliximab in patients with ulcerative colitis. *Gut* 2009;**58**(12):1612-9.

Regional variation in gene expression in the healthy colon is dysregulated in ulcerative colitis

C L Noble, A R Abbas, J Cornelius, C W Lees, G-T Ho, K Toy, Z Modrusan, N Pal, F Zhong, S Chalasani, H Clark, I D Arnott, I D Penman, J Satsangi and L Diehl

Gut 2008;57;1398-1405; originally published online 3 Jun 2008;
doi:10.1136/gut.2008.148395

Regional variation in gene expression in the healthy colon is dysregulated in ulcerative colitis

C L Noble,[1,2] A R Abbas,[3] J Cornelius,[2] C W Lees,[1] G-T Ho,[1] K Toy,[4] Z Modrusan,[4] N Pal,[5] F Zhong,[5] S Chalasani,[5] H Clark,[3] I D Arnott,[1] I D Penman,[1] J Satsangi,[1] L Diehl[2]

upplementary figures, tables methods are published e only at http://gut.bmj.content/vol57/issue10

trointestinal Unit, University inburgh, Western General ital, Edinburgh, UK; artment of Pathology, ntech, South San isco, USA; [3] Department of formatics, Genentech, n San Francisco, USA; roarray Department, ntech, South San isco, USA; [5] Core Labs, ntech, South San isco, USA

spondence to: Noble, Gastrointestinal Molecular Medicine e, Western General ital, Edinburgh, EH4 2XU, noblecolin@hotmail.com

nd LD are joint senior rs.

ed 29 April 2008
pted 19 May 2008
shed Online First
ne 2008

ABSTRACT

Objective: To investigate differential intestinal gene expression in patients with ulcerative colitis and in controls.

Design: Genome-wide expression study (41 058 expression sequence tags, 215 biopsies).

Setting: Western General Hospital, Edinburgh, UK, and Genentech, San Francisco, USA.

Patients: 67 patients with ulcerative colitis and 31 control subjects (23 normal subjects and 8 patients with inflamed non-inflammatory bowel disease biopsies).

Interventions: Paired endoscopic biopsies were taken from 5 specific anatomical locations for RNA extraction and histology. The Agilent microarray platform was used and confirmation of results was undertaken by real time polymerase chain reaction and immunohistochemistry.

Results: In healthy control biopsies, cluster analysis showed differences in gene expression between the right and left colon. ($\chi^2 = 25.1$, $p<0.0001$). Developmental genes, homeobox protein A13 (HOXA13), ($p = 2.3\times10^{-16}$), HOXB13 ($p<1\times10^{-45}$), glioma-associated oncogene 1 (GLI1) ($p = 4.0\times10^{-24}$), and GLI3 ($p = 2.1\times10^{-28}$) primarily drove this separation. When all ulcerative colitis biopsies and control biopsies were compared, 143 sequences had a fold change of >1.5 in the ulcerative colitis biopsies ($0.01>p>10^{-45}$) and 54 sequences had a fold change of <−1.5 ($0.01>p>10^{-20}$). Differentially upregulated genes in ulcerative colitis included serum amyloid A1 (SAA1) ($p<10^{-45}$) the alpha defensins 5 and 6 (DEFA5 and 6) ($p = 0.00003$ and $p = 6.95\times10^{-7}$, respectively), matrix metalloproteinase 3 (MMP3) ($p = 5.6\times10^{-10}$) and MMP7 ($p = 2.3\times10^{-7}$). Increased DEFA5 and 6 expression was further characterised to Paneth cell metaplasia by immunohistochemistry and in situ hybridisation. Subanalysis of the inflammatory bowel disease 2 (IBD2) and IBD5 loci, and the ATP-binding cassette (ABC) transporter genes revealed a number of differentially regulated genes in the ulcerative colitis biopsies.

Conclusions: Key findings are the expression gradient in the healthy adult colon and the involvement of novel gene families, as well as established candidate genes in the pathogenesis of ulcerative colitis.

The inflammatory bowel diseases (IBDs), ulcerative colitis and Crohn's disease, are chronic relapsing inflammatory diseases of the gastrointestinal tract. Both diseases are complex clinical entities that occur in genetically susceptible individuals who are exposed to as yet poorly defined environmental stimuli.[1 2]

In recent years, the application of non-parametric linkage analyses and well-designed case–control association studies have led to the identification of a number of susceptibility genes or loci strongly associated with Crohn's disease and ulcerative colitis. In Crohn's disease, success has been especially evident since the application of genome-wide association studies that have implicated at least 10 novel loci, including the genes encoding interleukin 23R (IL23R), autophagy-related 16-like 1 (ATG16L1) and immunity-related GTPase family, M (IRGM).[3–6] Determinants implicated to date in ulcerative colitis include the IBD2 and IBD5 loci, the human leukoryte antigen (HLA) complex, and the multidrug-resistance 1 (MDR1) gene.[1 7] These studies all highlight important pathways involved in disease pathogenesis and the emerging challenge now is to move from gene identification to functional understanding.

Microarray expression analysis allow a comprehensive picture of gene expression at the tissue and cellular level, thus helping to understand the underlying patho-physiological processes.[8] Earlier studies using microarray platforms to interrogate resection specimens from patients with IBD identified a number of novel genes that were differentially regulated when diseased samples from patients with severe disease were compared to controls.[9–11]

Subsequently, the ability to use endoscopic pinch mucosal biopsies rather than resection samples has allowed investigators to microarray tissue from a larger range of patients encompassing those with less severe disease.[12–15] Whilst these studies have shown consistent differential expression of a number of genes of interest, the data generated also highlight the heterogeneity of different experimental microarray platforms and biopsy collection techniques. As microarray technology has improved still further, data sets have become even more consistent, and the degree of replication between individual arrays and across platforms has increased significantly.[16 17]

The aims of the current study were to use microarray gene expression analysis to investigate genome-wide expression in endoscopic mucosal biopsies taken at colonoscopy from 67 patients with ulcerative colitis and 31 controls. This represents the largest cohort described to date and particular care has been taken to characterise disease phenotype as well as the anatomical location sampled. Comparisons have been made between gene expression profiles in health and disease, as well as within ulcerative colitis. Thereby, the present data set has allowed us to gain valuable insight into gradients of gene expression along the healthy adult colon, as well as changes associated with disease. The data

le 1 Demographics of the patients with ulcerative colitis

	Ulcerative colitis
number of patients	67
/female	33/34
an age at diagnosis (years)	37
an duration of follow-up (years)	7.8
ase group	
diagnosis	8
scent disease	41
e disease	18
ase extent at time of endoscopy	
itis	15
sided colitis	27
sive colitis	25
r features	
nt smoker	6
y history of IBD	5
A therapy	40
costeroid therapy	10
unosuppressant therapy (AZA, 6MP, MTX)	11

SA, 5-aminosalicylate; AZA, azathioprine; IBD, inflammatory bowel disease; P, 6-mercaptopurine; MTX, methotrexate.

end current understanding of gene expression in health, and complementary to current studies of germline and somatic ation associated with ulcerative colitis.

THODS

ents and controls

y-seven patients with ulcerative colitis and 31 control ents undergoing colonoscopy were recruited (table 1). All ents with ulcerative colitis attended the clinic at the stern General Hospital, Edinburgh, and the diagnosis of rative colitis adhered to the criteria given by Lennard-es.[18] Phenotypic data were collected by interview and case-e review.

leven of the controls were men and 20 were women. The lian age was 43 years at the time of endoscopy. Six of the trols had normal colonoscopies for colon cancer screening, e controls had symptoms consistent with irritable bowel drome and had a normal colonoscopic investigation and en patients had a colonoscopy for another indication and ologically normal biopsies were obtained. Eight control ents had abnormal inflamed colonic biopsies (one pseudo-mbranous colitis, one diverticulitis, one amoebiasis, two roscopic colitis, one eosoinophilic infiltrate, two scattered phoid aggregates and a history of gastroenteritis). Written rmed consent was obtained from all patients.

e 2 The location, number and inflammation status of biopsies in ents with ulcerative colitis and in controls

	Patients with ulcerative colitis		Controls	
number of paired ies	139		76	
oved from analysis	10		3	
	Inflamed	**Non-inflamed**	**Inflamed**	**Non-inflamed**
inal ileum	0	4	1	5
nding colon	12	21	4	13
ending colon	15	20	6	17
oid colon biopsies	35	22	9	18

Biopsy collection

Paired biopsies were taken from each anatomical location (table 2). One biopsy was sent for histological examination and the other was snap frozen in liquid nitrogen for RNA extraction.

Each biopsy was graded histologically by an experienced gastrointestinal pathologist as having no evidence of inflammation, biopsies with evidence of chronic inflammation and predominately chronic inflammatory cell infiltrate or simply those with acute inflammation and an acute inflammatory cell infiltrate.

Microarray analysis

Total RNA was extracted from each biopsy using the micro total RNA isolation from animal tissues protocol (Qiagen, Valencia, California, USA). One microgram of total RNA was amplified using the Low RNA Input Fluorescent Linear Amplification protocol (Agilent Technologies, Palo Alto, California, USA). A T7 RNA polymerase single round of linear amplification was carried out to incorporate the cyanine-3 and cyanine-5 labels into cRNA. The cRNA was purified using the RNeasy Mini Kit (Qiagen). One mictrolitre of cRNA was quantified using the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, Delaware, USA). Seven hundred and fifty nanograms of Universal Human Reference (Stratagene, La Jolla, California, USA) cRNA labelled with cyanine-3 and 750 ng of the test sample cRNA labelled with cyanine-5 were fragmented for 30 min at 60°C before loading onto the Agilent Whole Human Genome microarrays which are annotated to represent 33 296 genes.

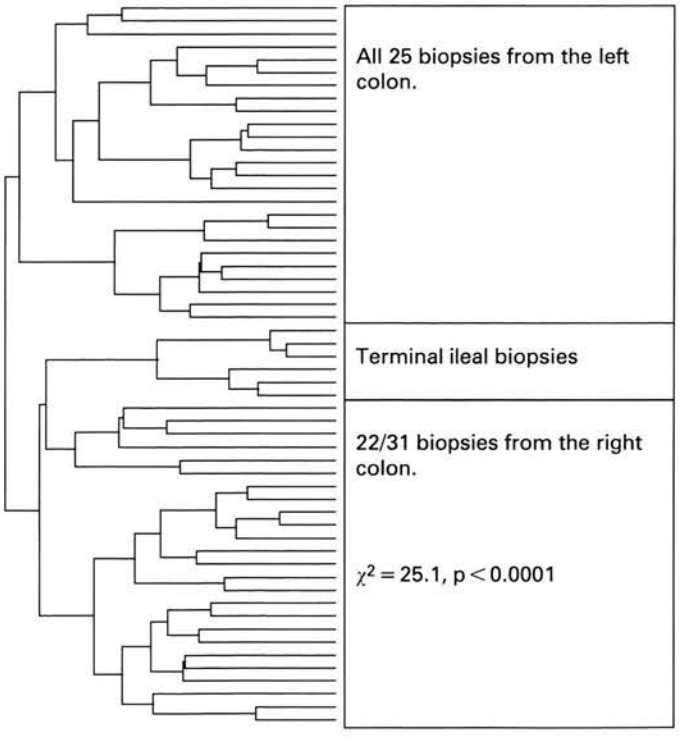


Figure 1 Histologically normal biopsies from control patients were analysed by unsupervised hierarchical clustering. Separation of the biopsies by anatomical location was observed and this separation was predominately driven by genes involved in the embryological development of the gastrointestinal tract.

The samples were hybridised for 18 h at 60°C with constant ation. Microarrays were washed, dried and scanned on the lent scanner according to the manufacturer's protocol. croarray image files were analysed using Agilent's Feature raction software version 7.5. The genes were normalised ng the Stratagene Universal Human Reference. The distribu- of log intensities for each sample was plotted and outlier ples (ie, greater than 2 standard deviations from the mean) e excluded from analysis. The whole data set is available at ne Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/ essed 18 July 2008)) accession: GSE11223.

, in situ hybridisation and immunohistochemistry

thods for real time polymerase chain reaction (PCR), in situ ridisation for defensin alpha 5, and immunohistochemistry rabbit anti-human lysozyme and rabbit anti-human defensin na 6 are provided in the supplementary methods section.

a analysis

croarray data were analysed using the Rosetta Resolver ware (Rosetta Inpharmatics, Seattle, WA, USA). Statistical ificance of the microarray data was determined by Student's aired t test. A value of $p<0.01$ and a fold change of greater ess than 1.5 were considered statistically significant. Fold nge data were calculated using the Rosetta Resolver ware. To correct for multiple hypothesis testing a q-value calculated for each tested feature to estimate significance in ns of the false discovery rate (FDR) rather than the false itive rate. From the q-values a FDR was calculated using the hod proposed by Storey and Tibshirani.[19] A FDR of less than was calculated for each of the presented analysis. Gene ology was analysed using Ingenuity software (Ingenuity tems, Mountain View, CA, USA). The Mann–Whitney U was used to analyse the real time PCR data. A value of 0.05 was considered significant.

RESULTS

Influence of anatomical location on gene expression in the healthy colon and terminal ileum

Unsupervised hierarchical clustering analysis using probes that had a fold change of greater or less than 1.3 were used to interrogate 56 histologically normal biopsies from control patients. Clear separation by anatomical location was observed; on one side of the dendrogram 25/25 biopsies were from the left colon (descending colon or sigmoid colon) whereas on the other side of the dendrogram 20/31 biopsies were from the ascending colon ($\chi^2 = 25.1$, $p<0.0001$) (fig 1). Biopsies from individual patients did not cluster together. The genes driving the differential expression between the right and left colon which were causing the observed clustering were predominately involved in the embryological development of the gastrointestinal tract: homeobox proteins HOXA13, fold change (FC), +4.93, $p = 2.3\times10^{-16}$), HOXB13 (FC, +16.96, $p<1\times10^{-45}$); glioma-associated oncogenes GLI1 (FC, +2.2, $p = 4.0\times10^{-24}$) and GLI3 (FC, +2.3, $p = 2.1\times10^{-28}$) were all upregulated in the left colon. Sixty-one sequences had a fold change of greater than 1.5 and 44 sequences had a fold change of less than 1.5 in the left colon in the control biopsies. In the ulcerative colitis biopsies when non-inflamed left and right colonic biopsies were compared 26 sequences had greater than a 1.5-fold increase in expression and 21 sequences had less than a −1.5-fold decrease in the left colon.

When gene ontology was compared between these two analyses alternative splicing genes were the most differentially regulated genes in both of the analyses ($p = 1.7\times10^{-41}$ and $p = 7.9\times10^{-19}$ for controls and ulcerative colitis, respectively). In the control analysis the next three most differentially regulated gene groups were nuclear protein function ($p = 4.3\times10^{-29}$), metal binding ($p = 4.7\times10^{-25}$) and membrane function ($p = 4.2\times10^{-20}$). In ulcerative colitis oxido-reductase function ($p = 1.4\times10^{-13}$) was the second most differentially expressed group followed by mitochondrial function ($p = 4.7\times10^{-25}$) and

e 3 Expression changes in the genes of interest

analysed	FC of all ulcerative colitis vs controls*	p Value	FC of non-inflamed ulcerative colitis sigmoid vs non-inflamed control sigmoid†	p Value	FC of inflamed ulcerative colitis sigmoid vs inflamed control sigmoid‡	p Value	FC of inflamed ulcerative colitis sigmoid vs non-inflamed sigmoid ulcerative colitis§	p value
	+8.18	$<10^{-45}$	+2.0	0.00024	+17.5	2.9×10^{-21}	+16.5	$<10^{-45}$
sin alpha 5	+3.25	0.00003	+1.02	0.89	+7.27	6.3×10^{-30}	+8.44	$<10^{-45}$
sin alpha 6	+2.18	6.95×10^{-7}	−1.09	0.34	+4.41	9.7×10^{-9}	+6.72	4.16×10^{-19}
A8	+3.50	2.3×10^{-17}	+1.21	0.19	+9.75	2.4×10^{-24}	+6.84	1.16×10^{-19}
A9	+3.06	4.1×10^{-13}	+1.05	0.16	+7.53	6.4×10^{-12}	+7.11	1.96×10^{-32}
3	+2.17	5.6×10^{-10}	−1.55	0.0088	+11.0	1.22×10^{-37}	+8.15	2.32×10^{-35}
7	+2.29	2.3×10^{-7}	+1.16	0.080	+7.31	4.9×10^{-24}	+5.53	1.01×10^{-23}
	+2.05	4.2×10^{-11}	+1.10	0.26	+6.36	9.27×10^{-17}	+7.24	8.42×10^{-19}
	+1.34	4.5×10^{-7}	+1.15	0.18	+1.50	0.0044	+1.54	0.00073
3	+8.02	1.1×10^{-17}	−1.30	0.20	+7.53	2.93×10^{-13}	+10.5	1×10^{-38}
0	+1.30	0.00011	+1.25	0.020	+1.79	0.00002	+2.36	4.68×10^{-11}
31	−1.32	0.00091	+1.10	0.40	−1.82	5.6×10^{-6}	−1.92	9.0×10^{-10}
DRB1	+1.03	0.88	−3.0	0.0010	+3.30	0.033	+2.67	0.0011
	−1.12	0.31	−2.73	2.7×10^{-10}	−1.15	0.61	+1.23	0.092

changes (FCs) and p values are shown in a number of different genes in four different experiments. Genes of interest were included in this table if significant consistent changes xpression were observed across more than one experiment.

9 ulcerative colitis biopsies and 73 control biopsies. †22 ulcerative colitis biopsies and 18 control biopsies. ‡35 ulcerative colitis biopsies and 8 control biopsies. §35 inflamed rative colitis biopsies and 22 non-inflamed ulcerative colitis biopsies.

e annotation: ABCB1, ABC transporter; CCL20, chemokine (C-C motif) ligand 20; HLA-DRB1, HLA class II histocompatibility antigen, DRB1; IL, interleukin; MMP, matrix alloproteinase; S100A8, calcium-binding protein A8; SAA1, serum amyloid A1; TLR4, toll like receptor 4; TNIP3, TNFAIP3 (A20), interacting protein 3; TSLP, thymic stromal phopoietin isoform 1.

alytic activity ($p = 2.7 \times 10^{-11}$). There was a 34.5% overlap ween the differentially expressed ontology groups in the trol and ulcerative colitis analysis.

lysis of expression in ulcerative colitis and control biopsies

ng unsupervised hierarchical clustering we were unable to erentiate between biopsies from patients with ulcerative tis and from controls patients. In addition no clustering based he inflammation status of the biopsies was observed. When all ulcerative colitis biopsies (129) and control biopsies (73) were pared, 143 sequence probes had a fold change of greater than in the ulcerative colitis biopsies ($0.01 > p > 10^{-45}$) and 54 uences had a fold change of less than 1.5 ($0.01 > p > 10^{-20}$) pplementary table 1). Serum amyloid A1 (SAA1) was the most egulated gene (FC, +8.18, $p < 10^{-45}$). Other notably upregu- d genes were the S100 calcium-binding proteins A8 (FC, +3.50, 2.3×10^{-17}), S100A9 (FC, +3.06, $p = 4.1 \times 10^{-13}$), the alpha nsins, alpha 5 (DEFA5) (FC, +3.25, $p = 0.00003$), alpha 6 FA6) (FC, +2.18, $p = 6.95 \times 10^{-7}$) and the matrix metallopro- ases MMP3 (FC, +2.17, $p = 5.6 \times 10^{-10}$) and MMP7 (FC, +2.29, 2.3×10^{-7}). The differential gene expression of a number of didate genes of interest is shown in table 3.

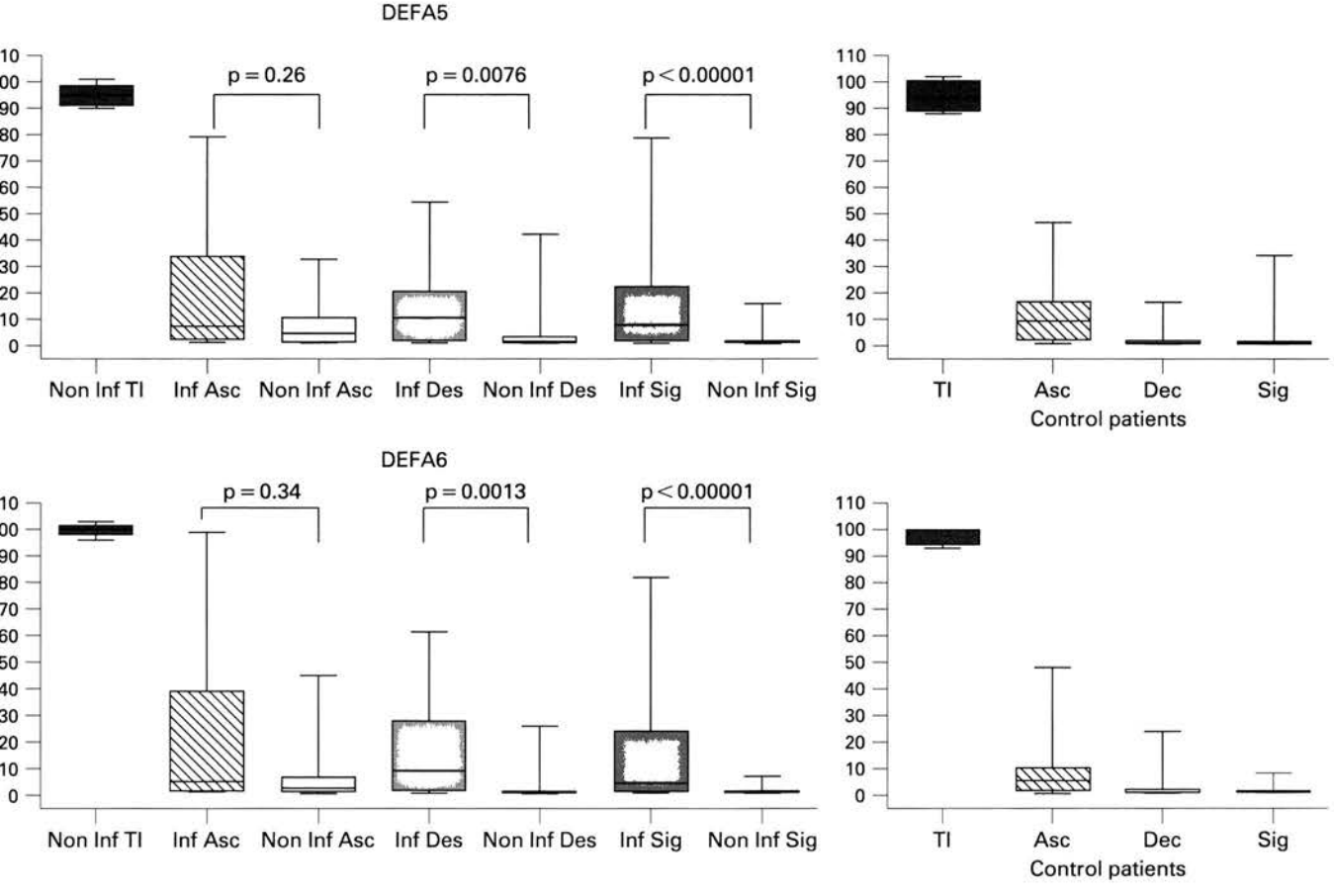
ene ontology analysis involving the genes differentially ressed between the ulcerative colitis and control biopsies wed a preponderance of differentially expressed genes were olved in immune response (48 genes out of a total of 679 es classified under immune response, $p = 2.1 \times 10^{-9}$, odds ratio (OR) 2.61, confidence interval (CI) 1.85 to 3.56) and response to wounding (30 genes out of a total of 359 genes classified under response to wounding, $p = 6.42 \times 10^{-9}$, OR 3.14, CI 2.09 to 4.53) when biological systems were considered.

Analysis of expression in sigmoid colon biopsies in patients with quiescent ulcerative colitis and non-inflamed control biopsies

To compare expression in biopsies without an acute inflammatory signal and to remove the effect of anatomical variation, 22 biopsies from the sigmoid colon with no histological evidence of inflammation from patients with ulcerative colitis were compared to 18 histologically normal control sigmoid colon biopsies. One hundred and two sequences had a fold change greater than 1.5 ($0.01 > p > 4.77 \times 10^{-13}$) and 84 sequences had a fold change of less than 1.5 ($0.01 > p > 1.8 \times 10^{-21}$) (supplementary table 2). Upregulated genes in the ulcerative colitis biopsies included defensin beta 14 (FC, +2.11, $p = 0.00002$) and SAA1 (FC, +2.01, $p = 0.00024$). Interesting genes that were downregulated included HLA class II histocompatibility antigen DRB1 (HLA-DRB1) (FC, −3.0, $p = 0.0010$) and thymic stromal lymphopoietin isoform (TSLP) (FC, −2.73, $p = 2.7 \times 10^{-10}$) (table 3).

Inflamed versus non-inflamed ulcerative colitis sigmoid colon biopsies

When expression signals were compared between 35 histologically inflamed and 22 non-inflamed sigmoid colon ulcerative colitis biopsies, 700 sequences had a fold change of greater than



re 2 Gene expression is shown as box-and-whisker plots. Each endoscopic biopsy has been separated by patient status, inflammation status and omical location. Significantly higher defensin alpha 5 (DEFA5) and DEFA6 expression was observed in the inflamed ulcerative colitis descending sigmoid colon biopsies when compared to the non-inflamed ulcerative colitis biopsies and the control biopsies. Asc, ascending colon; descending colon; Sig, sigmoid colon; TI, terminal ileum.

$(0.01>p>1\times10^{-45})$ and 518 sequences $(0.01>p>1\times10^{-45})$ a fold change of less than 1.5 in the inflamed biopsies pplementary table 3). Notably upregulated genes included A1 (FC, +16.5, $p=<10^{-45}$), TNFAIP3 interacting protein 3 IP3) (FC, +10.5, $p=1\times10^{-38}$), DEFA5 (FC, +8.44, $<10^{-45}$), DEFA6 (FC, +6.72, $p=4.16\times10^{-19}$) and regenerat-islet-derived 3 gamma (REG3γ) (FC, +6.99, $p=<10^{-45}$).

lysis of specific gene families: alpha defensins 5 and 6

en DEFA5 and DEFA6 were analysed expression in the mal controls and the non-inflamed ulcerative colitis biopsies similar across the different anatomical locations with there g high expression in the terminal ileum, and expression reasing as the biopsy location became more distal in the n (fig 2). In the acute and chronically inflamed ulcerative tis biopsies there was marked upregulation of DEFA5 and FA6 expression in the descending and sigmoid colon (table 3).

-binding cassette transporter family and the xenobiotic-scription regulators

ression patterns from probes representing 48 ATP-binding ette (ABC) transcriptional genes and their key mediators gnane X receptor (PXR), farnesoid X-activated receptor, and sterols receptor LXR-beta were analysed. When these genes e compared in all the ulcerative colitis and control biopsies, n genes were found to be significantly downregulated in the rative colitis samples when compared to the control ples: ABCA1 ($p=0.01$), ABCA8 ($p=0.0064$), ABCB1 0.00091), ABCC6 ($p=0.0050$), ABCB7 ($p=0.0068$), CF1 ($p=0.0005$) and ABCF2 ($p<0.00001$). Only one probe esenting ABCB2 was significantly upregulated in ulcerative tis ($p=0.0048$).

BCB1 expression was also significantly downregulated n inflamed ulcerative colitis sigmoid colon biopsies were pared to non-inflamed ulcerative colitis sigmoid biopsies -1.82, $p=5.6\times10^{-6}$) (table 3). No difference in the ression of PXR between ulcerative colitis and controls was rved in any of the analyses.

l time polymerase chain reaction analysis

ight genes implicated by microarray expression results, firmatory real time PCR analysis was undertaken in biopsies of patients form the original cohort following stratifying to represent a range of SAA1 and IL8 expression. Increased SAA1 expression in the inflamed ulcerative colitis sigmoid colon biopsies compared to the normal control sigmoid colon biopsies and the non-inflamed ulcerative colitis sigmoid colon biopsies ($p=0.041$ and $p=0.044$, respectively) was observed. Elevated IL8 expression was also confirmed in the inflamed ulcerative colitis sigmoid biopsies when compared to the control sigmoid biopsies ($p=0.031$) (supplementary fig 1).

Increased expression of DEFA5 and DEFA6 in the inflamed ulcerative colitis sigmoid colon biopsies when compared to the non-inflamed ulcerative colitis sigmoid colon biopsies ($p=0.0008$ and $p=0.0005$, respectively) and the control sigmoid colon biopsies ($p=0.0002$ and $p=0.0001$, respectively) was observed (supplementary fig 1). Increased expression in the inflamed ulcerative colitis sigmoid colon biopsies when compared to the non-inflamed ulcerative colitis sigmoid colon biopsies was also observed when MMP7 ($p=0.0005$), S100A8 ($p=0.0029$) and toll-like receptor 4 (TLR4) ($p=0.019$) were examined (supplementary fig 2).

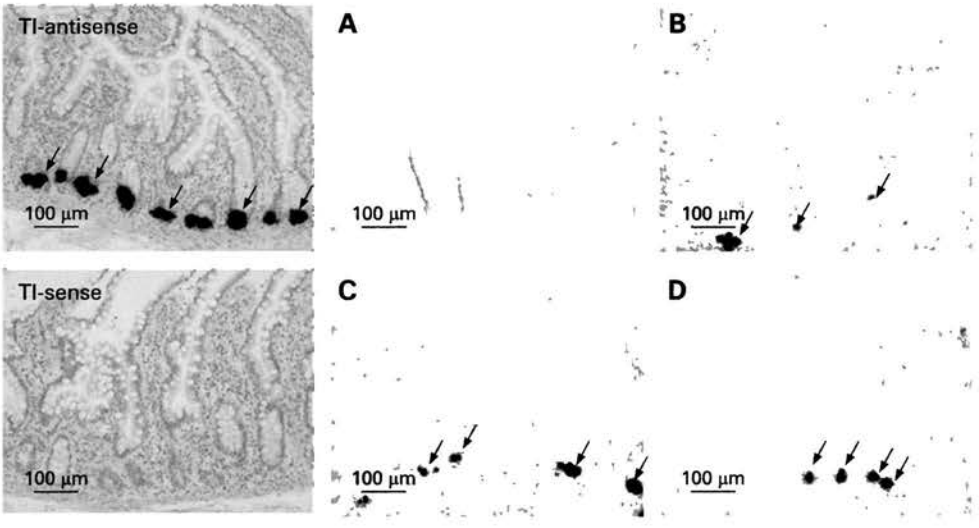
In situ hybridisation and immunohistochemistry

In situ hybridisation of the terminal ileal biopsies for DEFA5 showed strong hybridisation in the basal crypts consistent with Paneth cell location (fig 3). In the ulcerative colitis biopsies taken from the sigmoid colon strong, multifocal hybridisation in the basal crypt region of these biopsies was observed and this would be consistent with Paneth cell metaplasia. This was not observed in the non-inflamed control biopsies.

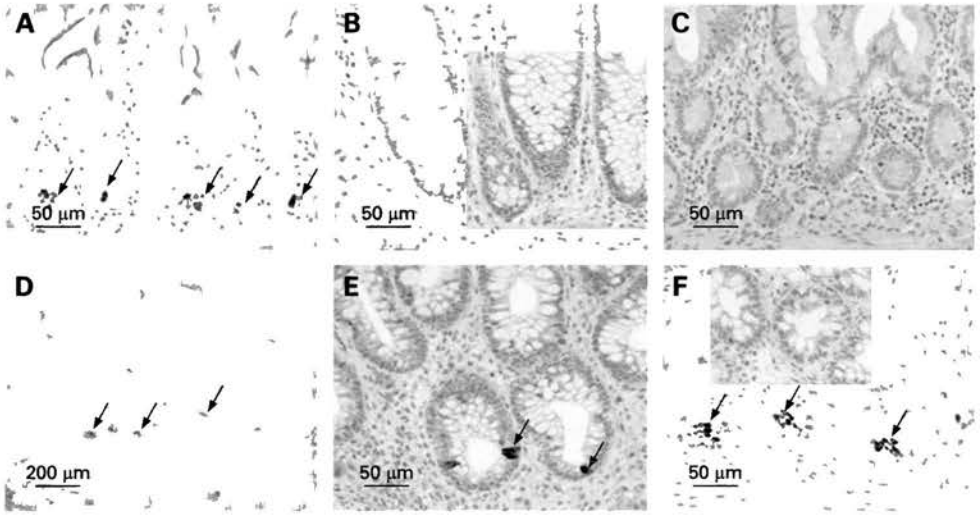
Immunohistochemistry for DEFA6 confirmed that in the sigmoid colon ulcerative colitis biopsies, staining was observed in the basal crypt region of these biopsies consistent with Paneth cell metaplasia. Again, this was not observed in the control biopsies (fig 4).

Expression of genes within the IBD2 locus

Using the markers defining the IBD2 locus we identified 526 Agilent probes representing genes or expressed sequence tags within this locus on chromosome 12. Twelve probes had a greater or less than a 1.5-fold change in expression with $p<0.01$ when expression of acute and chronically inflamed ulcerative colitis sigmoid colon biopsies were compared to non-inflamed ulcerative colitis sigmoid colon biopsies (table 4).



re 3 Terminal ileum (TI) upper I: antisense probe shows strong idisation in the basal crypts istent with Paneth cell location. ws indicate positive staining in crypt . Lower panel: no significant idisation with sense control probe. Sigmoid colon biopsy of a non-med control patient. (B,C,D) Strong, ifocal hybridisation in the basal crypt n of sigmoid colon biopsies from e patients with ulcerative colitis istent with Paneth cell metaplasia.



ıre 4 (A) Terminal ileal unohistochemistry shows positive ıing marked with arrows in the basal ts consistent with Paneth cell tion. (B,C) No significant staining was erved in two non-inflamed control ents. (D,E,F) Strong, multifocal ing in the basal crypt region of noid colon biopsies from three ents with ulcerative colitis consistent Paneth cell metaplasia.

ression of genes within the IBD5 locus

ent probes representing 11 genes within the IBD5 locus e identified and compared in healthy control, non-inflamed inflamed ulcerative colitis biopsies (table 5). Solute carrier ily 22A5 (SLC22A5) (OCTN2) was downregulated in rative colitis biopsies compared to controls (FC, −1.26 3.37×10^{-6}) and when inflamed ulcerative colitis sigmoid n biopsies were compared to non-inflamed ulcerative colitis noid colon biopsies (FC, −1.50, $p = 2.2\times10^{-6}$).

CUSSION

he present study we have taken particular care to address cerns currently being expressed about micoarray studies in ase. We initially documented the regional variation of gene ression in the healthy colon before undertaking comprehen- studies in ulcerative colitis. By documenting the anatomical tion for each biopsy, using appropriate inflammatory trols, as well as avoiding the pooling of samples, we have n able to remove a considerable amount of background ability that has hampered previous studies.[17 20–22] A further strength of our study has been the fact that real time PCR analysis consistently confirmed the significant changes in expression, substantially increasing the confidence associated with the interpretation of the data.

We have made a number of novel observations. In the healthy adult colon this is the first microarray study to show a gradient of expression of a number of genes. Genes involved in developmental pathways, the HOX family and the hedgehog signalling pathway, appeared to be the most differentially regulated along the anatomical length of the healthy colon. HOXA13 has been shown to play a crucial role in the development of the tail gut, and mutations in the gene result in urogenital abnormalities,[23] and interestingly it has been shown that HOXB13 expression is downregulated in colorectal tumours from the distal left colon.[24]

GLI1 is one of the major effector molecules of the hedgehog signalling pathway and the GLI1 gene lies within the IBD2 locus, strongly implicated in ulcerative colitis.[25–27] Data from the present study would also suggest that GLI1 is downregulated in inflamed ulcerative colitis biopsies compared to non-inflamed

Table 4 Gene expression in probes from the inflammatory bowel disease 2 (IBD2) locus, comparing inflamed ulcerative colitis sigmoid colon biopsies to non-inflamed ulcerative colitis sigmoid colon biopsies

Agilent probe	Gene symbol	FC of ulcerative colitis sigmoid inflamed vs non-inflamed*	p Value
A_23_P98876	SLC39A5	−1.52855	1.36×10^{-7}
A_24_P647146	HDAC7A	1.53876	3.15×10^{-16}
A_24_P941773	DKFZP586A0522	−2.27306	2.31×10^{-11}
A_24_P945113	ACVRL1	1.57956	6.33×10^{-9}
A_23_P128230	NR4A1	1.81844	0.00005
A_23_P331098	K5B	2.3845	0.0004
A_24_P246636	A_24_P246636	−1.69754	7.06×10^{-8}
A_23_P2233	SILV	1.51557	9.07×10^{-9}
A_23_P105251	GLI	−1.54447	7.73×10^{-9}
A_32_P3783	HMGA2	−1.94107	2.37×10^{-9}
A_23_P162300	IRAK3	1.7382	3.11×10^{-15}
A_32_P83256	IRAKM	1.70847	4.56×10^{-10}

Analysis of the 526 expression probes located within the IBD2 locus identified 12 probes that were significantly differentially regulated when the inflamed ulcerative colitis sigmoid colon biopsies were compared to the non-inflamed ulcerative colitis sigmoid colon biopsies.

*35 inflamed biopsies and 25 non-inflamed.

Gene annotation: A_24_P246636, unknown; ACVRL1, activin A receptor type II, like 1; DKFZP586A0522, methyltransferase-like protein 7A precursor; GLI, glioma-associated oncogene; HDAC7A, histone deacetylase 7A; HMGA2, high mobility group AT-hook 2; IRAK3, interleukin 1 receptor-associated kinase 3; IRAKM, interleukin 1 receptor-associated kinase M; KB5, keratin 5B; NR4A1, nuclear receptor subfamily 4, group A, member 1; SILV, silver homolog (mouse); SLC39A5, solute carrier family 39A5.

le 5 Fold expression changes in genes within the inflammatory bowel disease 5 (IBD5) locus in patients with ulcerative colitis and in controls

e analysed	FC of all ulcerative colitis vs controls*	p Value	FC of inflamed sigmoid vs non-inflamed sigmoid (ulcerative colitis)†	p Value	FC of non-inflamed ulcerative colitis sigmoid vs non-inflamed control sigmoid‡	p Value
	+1.00	0.96	+1.06	0.62	+1.14	0.40
	+1.12	0.057	−1.02	0.82	1.00	0.98
50	+1.02	0.20	−1.20	3.5×10^{-6}	+1.08	0.062
	+1.09	0.037	+1.10	0.17	+1.04	0.53
	+1.10	0.35	+1.12	2.9×10^{-6}	−1.01	0.62
2A5	−1.26	3.37×10^{-6}	−1.50	2.2×10^{-6}	+1.02	0.75
2A4	−1.18	0.11	−1.79	1.53×10^{-9}	+1.22	0.63
M4	+1.10	0.0056	+1.14	0.00039	+1.01	0.78
A2	−1.05	0.13	1.00	0.98	−1.05	0.28
	+1.10	0.056	+1.19	0.052	−1.04	0.63
	−1.01	0.39	+1.02	0.67	+1.01	0.75

l changes (FCs) in expression of genes within the IBD5 locus comparing controls and patients with ulcerative colitis, who have been stratified for the degree of inflammation erved in their sigmoid colon biopsies. Significant downregulation of the organic cation transporters SLC22A4 and SLC22A5 was observed when inflamed ulcerative colitis noid biopsies were compared to non-inflamed ulcerative colitis sigmoid biopsies.
9 ulcerative colitis and 73 controls. †35 inflamed sigmoid and 22 non-inflamed. ‡22 ulcerative colitis and 18 controls.
e annotation: CSF2, colony stimulating factor 2; IL, interleukin; IRF1, interferon regulatory factor 1; P4HA2, prolyl 4-hydroxylase subunit alpha-2 precursor; PDLIM4, PDZ and LIM ain 4; RAD50, DNA repair protein RAD50; SLC22, solute carrier family 22.

psies from patients with ulcerative colitis. These data add her weight to data from Cambridge and our own unit wing an association between mutations in the GLI1 gene ulcerative colitis.[28 29]

Vith regards to the observed gradient of expression in the lthy adult colon, our data contrast with data from Costello *al*[12] and Wu *et al*[15] where no significant differences in ression patterns were observed when comparing biopsies n caecum, transverse colon, descending colon and sigmoid on. The observed differences in these data sets may be lained by the fact that we have considered only non-amed healthy controls in these studies and have not pooled lthy and diseased data. When anatomical variation in the on was compared between patients with ulcerative colitis controls there was only a 34.5% homology between the erentially expressed gene ontology groups. This difference y be explained by environmental factors in the ulcerative cis colon such as microbial dysbiosis.

n ulcerative colitis, we demonstrated dysregulation of genes olved in innate immunity, notably the alpha defensins 5 and ogether with other pathways currently implicated in ammatory bowel disease. In our data set high levels of a defensin 5 and 6 expression were observed in the terminal biopsies of non-inflamed controls and patients with rative colitis. Levels of expression in these patients fell as location that the biopsies were retrieved from became more al in the colon: the ascending colon, descending colon and noid colon. However, in the inflamed ulcerative colitis psies increased expression of both alpha defensins 5 and 6 observed at each anatomical location. Lawrance and eagues[11] also noted that the defensins alpha 5 and 6 were egulated in patients with ulcerative colitis compared to trols, although RNA was extracted from surgical resections no details about the anatomical location of these specimens e given.

ecent data published by Varnat and colleagues[30] have gested that peroxisome proliferator activated receptor β ARβ) negatively regulates Paneth cell differentiation by ynregulating the expression of Indian hedgehog, another of major effector molecules in the hedgehog signalling path-y. Immunohistochemistry and in situ hybridisation have wn that this is largely mediated by Paneth cell metaplasia. Given our present data now implicating germline GLI1 variation in disease susceptibility, and the regional variation in health, we speculate that in patients with ulcerative colitis, further, as yet undetermined, defects in the hedgehog signalling pathway may result in unregulated Paneth cell differentiation, Paneth cell metaplasia, increased alpha defensin 5 and 6 expression, and mucosal inflammation.

It is of interest that many but not all of our results are broadly in line with two of the landmark microarray papers in IBD. Consistent with data from Lawrance and colleagues[11] we have shown upregulation of S100A8 and A9, and the alpha defensins 5 and 6 in ulcerative colitis. Dieckgraefe and colleagues[9] observed upregulation of a number of the MMP genes, again seen in our dataset. Another consistent finding was the upregulated expression of members of the REG family in the colon of patients with ulcerative colitis, probably as a result of Paneth cell metaplasia.[31]

The downregulation of ABCB1 in our dataset is of significant interest, and consistent with earlier microarray data from Langmann *et al*,[13] Dieckgraefe *et al*,[9] Lawrance *et al*,[11] and Wu *et al*.[15] It is pertinent that when the entire class of proteins sharing homology with ABCB1 were analysed, a further six out of 48 ABC transporters were significantly dysregulated in ulcerative colitis, including ABCA1, ABCA8, ABCC6, ABCB7, ABCF1 and ABCF2. An important role of the ABC transporters in the aetiopathogenesis of ulcerative colitis seems likely, supported by these consistent microarray data, association with germline MDR1 variability, and animal data to date.

However, and in contrast to data produced by Langmann *et al*,[13] we did not observe any changes in expression of the transcriptional regulator pregnane-X receptor. These negative data are consistent with genetic studies carried out in the IBD population in Edinburgh; using a haplotype tagging approach, there was an association between the ABCB1 gene and ulcerative colitis,[32] but no association between the pregnane-X receptor and ulcerative colitis.[33] Aspects of study design and patient recruitment may explain the differences observed between our data and those of Langmann and colleagues.

When novel genes were considered, of particular note was the differential expression of the poorly characterised gene TNFAIP3-interacting protein 3 (TNIP3) which is a nuclear or cytoplasmic protein with three coiled domains that was first

led down from a yeast-2 hybrid scan of TNFAIP3 (A20).[34] e protein has been found to be expressed in macrophages and upregulated by infection with *Listeria*. Further to this, FAIP3 was recently identified as the closest gene to 753394, a "second tier" hit from the UK genome-wide scan Crohn's disease, focusing further interest on this nuclear or κB dependent signalling pathway.[3]

The tight linkage disequilibrium spanning the IBD5 linkage erval has limited genetic studies of this region.[35–39] Consistent vnregulation of the organic cation transporter SLC22A5 CTN2) was observed in ulcerative colitis biopsies compared controls and in inflamed ulcerative colitis biopsies compared non-inflamed biopsies. SL22A4 (OCTN1) was also down-ulated in inflamed ulcerative colitis biopsies and these vocative data add weight to the hypothesis that decreased ression of these genes may after all be involved in the hogenesis of IBD.[40] However, we also noted expression of erferon regulatory factor 1 (IRF-1) and PDZ and LIM domain PDLIM4), both plausible candidates within IBD5 to be regulated, emphasising the uncertainties pertaining to this us at present.

n conclusion, these data provide a rigorously characterised ression profile of the whole genome in the terminal ileum colon of patients with ulcerative colitis and controls. The dies provide new insights into regional variation of gene ression in the healthy colon, and also considerably extend vious studies in ulcerative colitis. These data also identify a mber of key regulators of intestinal inflammation worthy of ther study. As further data from genome-wide scanning erge in this and other complex diseases, access to these data the ability to study expression, function and germline lation in parallel will become all the more necessary.

peting interests: None.

cs approval: Approval for this study was granted by the Lothian Research Ethics mittee: REC 04/S1103/22.

ERENCES

Cho JH, Weaver CT. The genetics of inflammatory bowel disease. *Gastroenterology* 2007;**133**:1327–39.

Strober W, Fuss I, Mannon P. The fundamental basis of inflammatory bowel disease. *J Clin Invest* 2007;**117**:514–21.

The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;**447**:661–78.

Duerr RH, Taylor KD, Brant SR, *et al*. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* 2006;**314**:1461–3.

Libioulle C, Louis E, Hansoul S, *et al*. Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLoS Genet* 2007;**3**:e58.

Raelson JV, Little RD, Ruether A, *et al*. Genome-wide association study for Crohn's disease in the Quebec Founder Population identifies multiple validated disease loci. *Proc Natl Acad Sci U S A* 2007;**104**:14747–52.

Gaya DR, Russell RK, Nimmo ER, *et al*. New genes in inflammatory bowel disease: lessons for complex diseases? *Lancet* 2006;**367**:1271–84.

Stoughton RB. Applications of DNA microarrays in biology. *Annu Rev Biochem* 2005;**74**:53–82.

Dieckgraefe BK, Stenson WF, Korzenik JR, *et al*. Analysis of mucosal gene expression in inflammatory bowel disease by parallel oligonucleotide arrays. *Physiol Genomics* 2000;**4**:1–11.

Heller RA, Schena M, Chai A, *et al*. Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc Natl Acad Sci U S A* 1997;**94**:2150–5.

Lawrance IC, Fiocchi C, Chakravarti S. Ulcerative colitis and Crohn's disease: distinctive gene expression profiles and novel susceptibility candidate genes. *Hum Mol Genet* 2001;**10**:445–56.

Costello CM, Mah N, Hasler R, *et al*. Dissection of the inflammatory bowel disease transcriptome using genome-wide cDNA microarrays. *PLoS Med* 2005;**2**:e199.

Langmann T, Moehle C, Mauerer R, *et al*. Loss of detoxification in inflammatory bowel disease: dysregulation of pregnane X receptor target genes. *Gastroenterology* 2004;**127**:26–40.

Okahara S, Arimura Y, Yabana T, *et al*. Inflammatory gene signature in ulcerative colitis with cDNA macroarray analysis. *Aliment Pharmacol Ther* 2005;**21**:1091–7.

Wu F, Dassopoulos T, Cope L, *et al*. Genome-wide gene expression differences in Crohn's disease and ulcerative colitis from endoscopic pinch biopsies: insights into distinctive pathogenesis. *Inflamm Bowel Dis* 2007;**13**:807–21.

Patterson TA, Lobenhofer EK, Fulmer-Smentek SB, *et al*. Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project. *Nat Biotechnol* 2006;**24**:1140–50.

Shi L, Reid LH, Jones WD, *et al*. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 2006;**24**:1151–61.

Lennard-Jones JE. Classification of inflammatory bowel disease. *Scand J Gastroenterol Suppl* 1989;**170**:2–6.

Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 2003;**100**:9440–5.

Brazma A, Hingamp P, Quackenbush J, *et al*. Minimum information about a microarray experiment (MIAME) – toward standards for microarray data. *Nat Genet* 2001;**29**:365–71.

Ioannidis JP. Microarrays and molecular research: noise discovery? *Lancet* 2005;**365**:454–5.

Marshall E. Getting the noise out of gene arrays. *Science* 2004;**306**:630–1.

De Santa BP, Roberts DJ. Tail gut endoderm and gut/genitourinary/tail development: a new tissue-specific role for Hoxa13. *Development* 2002;**129**:551–61.

Jung C, Kim RS, Zhang H, *et al*. HOXB13 is downregulated in colorectal cancer to confer TCF4-mediated transactivation. *Br J Cancer* 2005;**92**:2233–9.

Lees C, Howie S, Sartor RB, *et al*. The hedgehog signalling pathway in the gastrointestinal tract: implications for development, homeostasis, and disease. *Gastroenterology* 2005;**129**:1696–710.

Parkes M, Barmada MM, Satsangi J, *et al*. The IBD2 locus shows linkage heterogeneity between ulcerative colitis and Crohn disease. *Am J Hum Genet* 2000;**67**:1605–10.

Satsangi J, Parkes M, Louis E, *et al*. Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. *Nat Genet* 1996;**14**:199–202.

Lees CW, Nimmo ER, Ho GT, *et al*. Analysis of the IBD2 locus: inherited variants in GLI-1, the major effectors of the hedgehog signalling pathway are strongly implicated in susceptibility to ulcerative colitis. *Gastroenterology* 2006;**130**:A52.

Tremelling M, John S, Berzuini C, *et al*. Positional candidate gene analysis in IBD2 confirms association between ulcerative colitis and GLI-1 but not IL23A. *Gastroenterology* 2007;**132**:A152.

Varnat F, Heggeler BB, Grisel P, *et al*. PPARbeta/delta regulates paneth cell differentiation via controlling the hedgehog signaling pathway. *Gastroenterology* 2006;**131**:538–53.

Ogawa H, Fukushima K, Naito H, *et al*. Increased expression of HIP/PAP and regenerating gene III in human inflammatory bowel disease and a murine bacterial reconstitution model. *Inflamm Bowel Dis* 2003;**9**:162–70.

Ho GT, Soranzo N, Nimmo ER, *et al*. ABCB1/MDR1 gene determines susceptibility and phenotype in ulcerative colitis: discrimination of critical variants using a gene-wide haplotype tagging approach. *Hum Mol Genet* 2006;**15**:797–805.

Ho GT, Soranzo N, Tate SK, *et al*. Lack of association of the pregnane X receptor (PXR/NR1I2) gene with inflammatory bowel disease: parallel allelic association study and gene wide haplotype analysis. *Gut* 2006;**55**:1676–7.

Staege H, Brauchlin A, Schoedon G, *et al*. Two novel genes FIND and LIND differentially expressed in deactivated and Listeria-infected human macrophages. *Immunogenetics* 2001;**53**:105–13.

Fisher SA, Hampe J, Onnie CM, *et al*. Direct or indirect association in a complex disease: the role of SLC22A4 and SLC22A5 functional variants in Crohn disease. *Hum Mutat* 2006;**27**:778–85.

Noble CL, Nimmo ER, Drummond H, *et al*. The contribution of OCTN1/2 variants within the IBD5 locus to disease susceptibility and severity in Crohn's disease. *Gastroenterology* 2005;**129**:1854–64.

Rioux JD, Daly MJ, Silverberg MS, *et al*. Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. *Nat Genet* 2001;**29**:223–8.

Russell RK, Drummond HE, Nimmo ER, *et al*. Analysis of the influence of OCTN1/2 variants within the IBD5 locus on disease susceptibility and growth indices in early onset inflammatory bowel disease. *Gut* 2006;**55**:1114–23.

Waller S, Tremelling M, Bredin F, *et al*. Evidence for association of OCTN genes and IBD5 with ulcerative colitis. *Gut* 2006;**55**:809–14.

Peltekova VD, Wintle RF, Rubin LA, *et al*. Functional variants of OCTN cation transporter genes are associated with Crohn disease. *Nat Genet* 2004;**36**:471–5.

The Contribution of OCTN1/2 Variants Within the *IBD5* Locus to Disease Susceptibility and Severity in Crohn's Disease

COLIN L. NOBLE,* ELAINE R. NIMMO,* HAZEL DRUMMOND,* GWO–TZER HO,* ALBERT TENESA,† LINDA SMITH,* NORMAN ANDERSON,* IAN D. R. ARNOTT,* and JACK SATSANGI*

*Gastrointestinal Unit and †MRC Human Genetics Unit, Western General Hospital, University of Edinburgh, Edinburgh, Scotland

See editorial on page 2106.

Background & Aims: **Recent data suggest that polymorphisms in the organic cation transporter (OCTN) genes OCTN1 (*SLC22A4*) and OCTN2 (*SLC22A5*) represent disease-causing mutations within the *IBD5* locus (chromosome 5q31). We investigated associations with disease susceptibility, phenotype, and evidence for epistasis with *CARD15* in 679 patients with Crohn's disease (CD) or ulcerative colitis (UC).** ***Methods:*** **A total of 374 patients with CD, 305 patients with UC, and 294 healthy controls (HCs) were studied. Genotyping for single nucleotide polymorphisms IGR2096, IGR2198, and IGR2230, OCTN1 variant (*SLC22A4* 1672C→T), and OCTN2 variant (*SLC22A5* −207G→C) was performed using the TaqMan system.** ***Results:*** **The *IBD5* OCTN1 and OCTN2 polymorphisms were in strong linkage disequilibrium (D', >0.959). IGR2198 variant allele frequency (49.1% vs 40.8%; $P = .0046$) and homozygosity (21% vs 14.8%; $P = .044$) were associated with CD versus HCs. Variant allelic frequency of OCTN1 (53.6% vs 48%; $P = .0008$) and OCTN2 (56.1% vs 48.4%; $P = .092$) polymorphisms and homozygosity for the OCTN1/2-TC haplotype (28.4% vs 16%; $P = .0042$) were associated with CD versus HCs. IGR2198 homozygosity and TC homozygosity were associated with stricturing/penetrating disease at follow-up ($P = .011$ and $P = .011$, respectively) and disease progression ($P = .038$ and $P = .049$, respectively) on univariate analysis and with need for surgery on multivariate analysis ($P = .016$ and $P = .004$, respectively). In the absence of the *IBD5* risk haplotype, no association of OCTN1/2 variants with CD was detected. No associations were seen with UC.** ***Conclusions:*** **The *IBD5* locus influences susceptibility, progression, and need for surgery in CD. However, the contribution of OCTN1/2 variants is not independent of the *IBD5* haplotype; a causative role for these genes remains plausible but is not yet proven. Further genetic, functional, and expression data are now required.**

The inflammatory bowel diseases (IBDs), Crohn's disease (CD) and ulcerative colitis (UC), are common causes of gastrointestinal morbidity in the Western world. The incidence of early-onset disease continues to increase in northern Europe, notably in Scotland and Scandinavia.[1–4] Epidemiologic, molecular, and clinical studies have proven that genetic susceptibility combined with environmental interaction are central to the pathogenesis of IBD.[5]

Genome-wide scanning has identified susceptibility loci for CD on chromosomes 1,[6] 5 (*IBD5*),[7–9] 6 (*IBD3*; *HLA*),[8,10] 12 (*IBD2*),[11] 14 (*IBD4*),[7,12] 16 (*IBD1*),[11,13] and 19 (*IBD6*).[8] The most consistently replicated CD susceptibility locus is located on chromosome 16 (*IBD1*), and the susceptibility gene has been identified as the *NOD2/CARD15* gene.[14,15] *NOD2/CARD15* contains a caspase recruitment domain (CARD) that is linked to a nucleotide-binding domain and a leucine-rich repeat region. *NOD2/CARD15* functions as an intracellular sensor of muramyl dipeptide, a highly conserved peptidoglycan motif common to many intraluminal bacteria.[16,17] Watanabe et al suggested that *NOD2/CARD15*$^{-/-}$ mice lose negative control of Toll-like receptor 2–mediated activation of nuclear factor κB, potentially offering an explanation for the Th1 phenotype characteristic of CD.[18] However, recently published studies do not provide support for *NOD2/CARD15* interaction with the Toll-like receptor 2 pathway, and these data emphasize the complexity of *NOD2/CARD15* activation.[19,20]

Two single nucleotide polymorphisms (SNPs) (Gly908Arg and Arg702Trp) and a frameshift mutation (Leu1007fsinsC) induce structural changes in the leucine-rich region of *NOD2/CARD15* and have been associated with CD. Reported *NOD2/CARD15* carriage

Abbreviations used in this paper: HC, healthy control; OCTN, organic cation transporter; SNP, single nucleotide polymorphism.

0016-5085/05/$30.00
doi:10.1053/j.gastro.2005.09.025

ites in CD vary between 0 and 50.0%, with highest ites seen in central European populations,[14,21] while nutations are absent in Japanese and Chinese series.[22,23] ower frequencies have been reported from Finland and cotland, suggesting heterogeneity within Europe.[24,25] *OD2/CARD15* polymorphisms have consistently been ssociated with a younger age of onset of CD, ileal isease, and fibrostenosing disease.[26]

The *IBD5* locus on chromosome 5q31-33 was origially identified in 1999 as conferring susceptibility for D.[7] This finding was replicated using linkage disequibrium mapping in the Canadian CD population and rther delineated to 5q31.[8] Fine mapping of this area as identified a single, highly conserved 250-kilobase aplotype of 11 SNPs spanning a cytokine gene cluster hat is associated with CD.[9] Further high-resolution nalysis of the 5q31 region using 103 SNPs revealed 11 screte haplotype blocks that measure tens of kilobases length, have limited diversity, and are punctuated by tes of recombination.[27] Genotype-phenotype analysis as suggested an association between the *IBD5* locus, erianal disease,[28] and early onset of disease,[29] but evience for epistasis between *IBD5* and *NOD2/CARD15* as been inconsistent.[9,28–30]

More recently, data have suggested that mutations of genes within the *IBD5* region, the organic cation ansporters OCTN1 (also known as *SLC22A4*) and CTN2 (*SLC22A5*), may be independently associated ith CD.[31] The construction of a 2-allele risk haplotype CTN1/2-TC (*SLC22A4* exon 9 1672C→T and *LC22A5* promoter −207G→C) was reported to be sociated with CD in patients who lacked the extended *BD5* risk haplotype. Individual allelic data were not ven.

OCTN1 is a 551–amino acid protein that is rongly expressed in the kidney, trachea, bone marow, and, to a lesser extent, small bowel and has been haracterized as a carnitine transporter.[32] An intronic NP (rs2268277) in a Runt-related transcription factor 1 (RUNX1) binding site of OCTN1 has been sociated with susceptibility to rheumatoid arthritis the Japanese population.[33] OCTN2 is a 557–amino cid protein that is 75.8% homologous to OCTN1, nd functional studies have shown it to be a highffinity sodium carnitine transporter that is expressed the kidney, smooth muscle, and heart tissue.[34] eltekova et al suggested that the OCTN1 variant 672C→T) alters its function in fibroblasts in vitro, ith variant forms having less affinity for carnitine nd a greater affinity for tetraethyl ammonium and me xenobiotics, and that the OCTN2 variant −207G→C) disrupted a heat-shock transcription factor binding site in fibroblasts in vitro.[31] No data demonstrating the altered function or expression of OCTN1/2 gene products in CD have yet been provided.

The provocative association data from Peltekova et al have yet to be replicated in an independent population, and epistasis with *NOD2/CARD15* has not thus far been assessed without the Canadian population. Phenotypic data from the same Canadian cohort of patients with CD have suggested a correlation between OCTN1/2 variants and ileal disease, and this correlation becomes substantially stronger when carriage of *NOD/CARD15* variants is taken into account.[35] However, the phenotypic description of this cohort does not allow for analysis of disease behavior, progression, or indeed proximal small bowel disease.

In the present study, we assessed the contribution of the OCTN1 and OCTN2 polymorphisms implicated by Peltekova et al in determining genetic susceptibility in CD and UC, specifically addressing whether these OCTN1/2 variants have an association with susceptibility to IBD that is independent of other markers within the extended *IBD5* linkage interval. We also investigated whether these polymorphisms are associated with specific disease phenotype characteristics or progression in our rigorously defined IBD population and assessed gene-gene interactions with established *NOD2/CARD15* mutations.

Materials and Methods

A total of 679 patients with well-characterized IBD and 294 healthy controls (HCs) were recruited. All patients with IBD attended the clinic at Western General Hospital (Edinburgh, Scotland). This is a tertiary referral center for IBD in southeastern Scotland. The IBD group comprised 374 patients with CD and 305 patients with UC. The diagnosis of IBD adhered to the criteria of Lennard-Jones.[36] Age at diagnosis, location, and behavior were classified according to the Vienna classification.[37] Phenotypic data were collected by patient questionnaire, interview, and case note review and were composed of demographics, date of onset and diagnosis, disease location, disease behavior, progression, extraintestinal manifestations, surgical operations, smoking history, joint symptoms, family history, and ethnicity. Written informed consent was obtained from all patients. The Medicine and Oncology Subcommittee of the Lothian Local Research Ethics Committee approved the study protocol (LREC 2000/4/192).

Demographics

CD group. The demographics of the patients recruited are shown in Table 1. The duration of follow-up was defined as the time from diagnosis to the time of most

able 1. Demographics and Clinical Features of the CD, UC, and Control Groups

	CD	UC	Controls
otal no.	374	305	294
ex (male/female)	181/193	171/134	143/151
edian age at diagnosis, *y* (interquartile range)	27.8 (20.9–40.4)	34 (25–50)	39 (27–52)
edian duration of follow-up, *y* (interquartile range)	11.8 (6.5–20.2)	7.5 (3.35–13.4)	
on-Jewish White (%)	98.7	98.3	
ge at diagnosis, A1 (younger than 40 years)/ A2 (older than 40)	72%/28%		
ocation at diagnosis (%)			
Ileal disease (L1)	125 (34)	Proctitis: 105 (34.5)	
Colonic disease (L2)	139 (38)	Left-sided colitis: 116 (37)	
Ileal and colonic disease (L3)	58 (16)	Extensive colitis: 84 (27.5)	
Upper gastrointestinal disease (L4)	30 (8)		
Oral CD	6 (2)		
Perianal disease	75 (21)		
sease behavior at diagnosis (%)			
Inflammatory (Vienna B1)	258 (74)		
Stricturing (Vienna B2)	32 (9)		
Penetrating (Vienna B3)	61 (17)		
sease behavior at follow-up (%)			
Inflammatory (B1)	128 (34)		
Stricturing (B2)	70 (19)		
Penetrating (B3)	168 (48)		
sease progression			
progression from inflammatory (% of inflammatory at diagnosis)	126 (49)		
Inflammatory to structuring or penetrating (%)	132 (51)		
Surgery for luminal complications of CD (%)[a]	237 (63)		

OTE. Disease behavior at follow-up was defined using the Vienna classification when the patient was last clinically evaluated. Full phenotypic ta were available on 94% of the patients with CD at diagnosis and 98% at follow-up. In assessing disease progression, patients were grouped those who remained as having inflammatory (B1), nonprogressive disease and those whose disease had progressed to stricturing (B2) or enetrating (B3) disease.

xamination under anesthesia or drainage procedures for perianal sepsis were excluded.

cent clinic review. Disease behavior at follow-up was efined using the Vienna classification of behavior at the me point when the disease was last clinically evaluated adiologically, endoscopically, or at surgery).[24,37] In assessg disease progression, patients who had Vienna B1 (inammatory) CD at diagnosis were identified and disease ehavior was compared with behavior at most recent follow-. Patients were grouped as those who remained classified having inflammatory (B1) disease at follow-up (nonpronessive disease), and those whose disease had progressed to ricturing (B2) or penetrating (B3) disease. Ninety-seven atients (26%) in the CD group were current smokers, and (21%) had a first-degree relative with a family history of D.

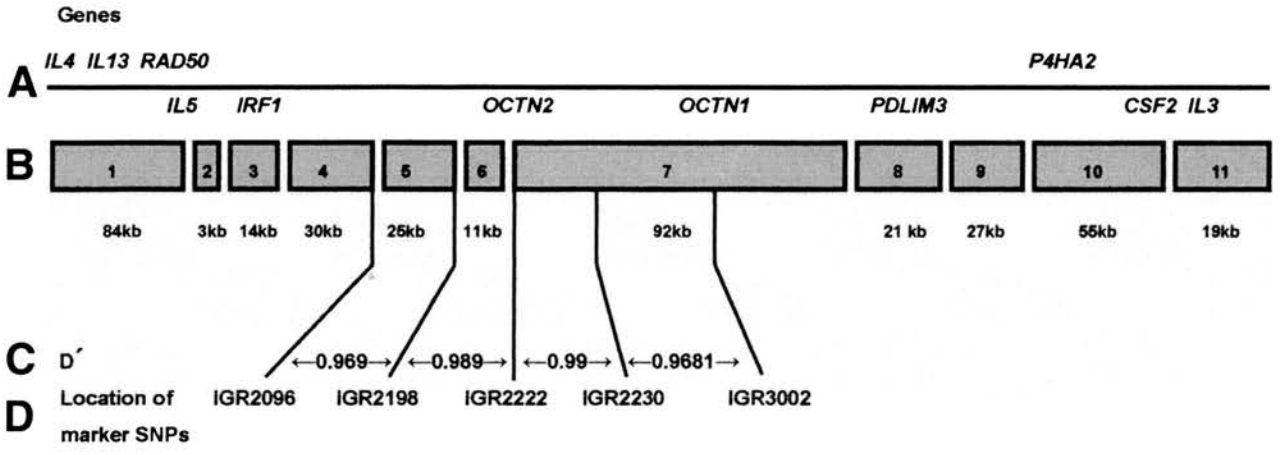
UC group. Clinical details were available regarding e entire UC group. The group consisted of 171 men and 134 omen with a median age at diagnosis of 34 years (Table 1). isease extent was recorded at diagnosis.

Control group. The control group comprised 294 ealthy subjects: blood donors (n = 163) recruited from utheastern Scotland and healthy control subjects (n = 131). here were 143 men and 151 women with a median age of 39 ars (interquartile range, 27–52 years).

Genotyping

Genomic DNA was extracted from peripheral venous blood by a modified salting-out technique[38] and resuspended in 1× Tris-EDTA buffer (10 mmol/L Tris [pH 8.0] and 1 mmol/L EDTA [pH 8.0]) at a final concentration of 100 ng/μL. To examine the relative contribution of the OCTN1/2 variants relative to the extended *IBD5* haplotype, 3 SNPs were genotyped: IGR2096 (which lies within haplotype block 4, as defined by Daly et al[27]) (Figure 1), IGR2198 (within haplotype block 5), and IGR2230 (within haplotype block 7). The rs1050152 polymorphism of the OCTN1 gene (*SLC22A4* exon 9 1672C→T, IGR3002) and the rs26313667 (*SLC22A5* promoter −207G→C, IGR2222) polymorphism of the OCTN2 gene were typed. Details of these variants and of the TaqMan primers are provided in Supplementary Table 1 (see supplemental material online at www.mmc.med.ed.ac.uk/gi/supdata.shtml) and Figure 1.

Patients with CD and HCs were typed for polymorphisms of the *NOD2/CARD15* gene (R702W, G908R, and 1007fsinsC) using previously described methods.[24] All genotyping except the R702W polymorphism was performed using the TaqMan system (7900HT sequence detection system; Applied Biosystems, Foster



gure 1. Haplotype structure and linkage disequilibrium across the *IBD5* region. The *IBD5* locus (5q31) with the high-resolution haplotype ructure as reported by Daly et al.[27] (*A*) Candidate genes above the relevant haplotype blocks. Genes above the line are transcribed from left right, and those below the line are transcribed from right to left. (*B*) Eleven blocks numbered 1–11 (between 3 and 92 kilobases), each of nited genetic diversity, are punctuated by sites of recombination. (*C*) D′ scores are shown to demonstrate the tight linkage disequilibrium tween the SNPs that were analyzed. The lowest D′ (0.959) was observed between IGR2096 and the OCTN2 variant (−207G→C). (*D*) Location the SNPs that were analyzed, with IGR2222 representing the OCTN2 variant (−207G→C) and IGR3002 representing the OCTN1 variant 672C→T). IGR2078, which was used by Peltekova et al to represent the extended *IBD5* haplotype, is located in block 4.[31]

ty, CA). R702W genotyping was performed by restriction igment length polymorphism polymerase chain reaction. Re- riction digestion was performed using 1 U *Msp* É at 37°C ernight and polymerase chain reaction fragments run on 4% uSieve 3:1 agarose gels (Cambrex Bio Science, Nottingham, d, Nottingham, UK). These were stained with ethidium bro- ide and viewed under UV light. An image was recorded gitally.

Data Analysis

Each SNP was analyzed for association with IBD erall, CD, UC, and disease phenotype. Allele frequencies d carrier frequencies were determined for all polymor- isms. Each of the variants studied was shown to be in ardy–Weinberg equilibrium in patients and in controls. enotype-phenotype associations were analyzed by χ^2 test ing the Minitab statistical software package (Minitab Ltd, oventry, England). Linkage disequilibrium and haplotype alysis were investigated using cocaphase (Rosland Frank- n Centre for genomics research; http://www.hgmp.mrc.a- uk). Evidence for epistasis among the OCTN1/2 allelic riants, the OCTN1/2-TC haplotype, and *NOD2/CARD15* riants was investigated by comparing allelic frequencies the individual OCTN1/2 variants, together with ho- ozygosity for the OCTN1/2-TC haplotype between the bgroups of patients with and without *NOD2/CARD15* riants by χ^2 analysis.

To identify significant independent variables associated th genotype, multiple logistic regression analysis was rformed using the Minitab statistical software package Minitab Ltd). The population-attributable risk percentage as defined as the excess rate of disease in individuals with mutation compared with those without. This was esti- ated by the method of Schlesselman,[39] which estimates population-attributable risk percentage as equal to attributable risk as a function of the prevalence in the exposed population divided by incidence of IBD in the population. To calculate this, the prevalence of CD was estimated at 100/100,000 and the frequency of all alleles in the control population was assumed to reflect that of the general population.

Results

Linkage Disequilibrium Across the *IBD5* Locus

In the CD population, strong linkage disequilibrium was observed between the SNPs within the extended *IBD5* haplotype as defined by Daly et al[27]: IGR2096 (block 4), IGR2198 (block 5), OCTN2 (*SLC22A5* promoter −207G→C) (block 7), IGR2230 (block 7), and OCTN1 (*SLC22A4* exon 9 1672C→T) (block 7) (Figure 1). Pairwise D′ values >0.959 between each SNP confirmed the difficulty of showing an independent effect of OCTN1/2 variants and *IBD5*. In the detailed analyses of the *IBD5* contribution to disease phenotype presented in this report, data are presented for the IGR2198 marker, representative of the 3-locus haplotype (IGR2096, IGR2198, and IGR2230). All data are available for each marker SNP on request.

Disease Susceptibility

Analysis of SNPs representing the extended *IBD5* haplotype (IGR2096, IGR2198, and IGR2230). Variant allelic frequencies differed significantly between patients with CD and HCs for each of the 3 *IBD5* SNPs studied

able 2. *IBD5* Variant Allele Frequency, Carriage Rates, and Homozygote Frequency in Patients with CD and Controls

SNP/haplotype examined	Control frequency (%)	CD frequency (%)	*P*	Odds ratio (confidence interval)
ariant allelic frequency				
IGR2096	41.9	49.4	.03	1.42 (1.1–1.8)
IGR2198	40.8	49.1	.0046	1.40 (1.1–1.6)
OCTN2 (−207 G→C)	48.4	56.1	.0092	1.43 (1.2–1.8)
IGR2230	47.4	54.9	.011	1.35 (1.1–1.7)
OCTN1 (1672C→T)	42.9	53.6	.0008	1.48 (1.2–1.9)
ariant carriage rates				
IGR2096	173 (68.7)	264 (79)	.0034	1.75 (1.2–2.6)
IGR2198	171 (67)	258 (77)	.011	1.61 (1.2–2.3)
OCTN2 (−207 G→C)	187 (74)	280 (81)	.034	1.52 (1.0–2.2)
IGR2230	185 (73.7)	261 (80)	.047	1.5 (1.1–2.23)
OCTN1 (1672C→T)	178 (68.7)	267 (80)	.0024	1.77 (1.2–2.6)
TC haplotype	170 (69)	264 (80)	.0016	1.8 (1.3–2.7)
IBD5 haplotype[a]	166 (63.6)	246 (73)	.016	1.5 (1.1–2.2)
ariant homozygote frequency				
IGR2096	38 (15.1)	73 (21.4)	.036	1.58 (1.0–2.4)
IGR2198	38 (14.8)	70 (21)	.044	1.56 (1.0–2.4)
OCTN2 (−207 G→C)	55 (22)	106 (31)	.0155	1.59 (1.1–2.3)
IGR2230	47 (18.6)	94 (29)	.0038	1.79 (1.2–2.6)
OCTN1 (1672C→T)	44 (17)	89 (25.6)	.011	1.69 (1.1–2.5)
TC haplotype	40 (16)	86 (28.4)	.0042	1.83 (1.2–2.8)
IBD5 haplotype[a]	38 (14.6)	68 (20.5)	.071	1.48 (1.0–2.3)

OTE. Three SNPs were used to define the extended *IBD5* locus (IGR2096, IGR2198, and IGR2230) (Figure 1). Each was independently ssociated with susceptibility to CD when allelic frequencies, carriage rates, and homozygosity were analyzed. OCTN2 variant (−207G→C) and CTN1 variant (1672C→T) were also independently associated with susceptibility to CD when allelic frequencies, carriage rates, and omozygosity were analyzed. The 2-allele risk haplotype OCTN1/2-TC haplotype was associated with susceptibility to CD when carriage rates and omozygosity rates were analyzed.

2-allele risk haplotype using the *IBD5* marker SNPs IGR2198 and IGR2230 is also illustrated for comparison with the OCTN1/2-TC haplotype.

GR2096, 49.4% CD vs 41.9% HC [P = .03]; GR2198, 49.1% CD vs 40.8% HC [P = .0046]; GR2230, 54.9% CD vs 47.4% HC [P = .011]) (Table . Carriage rates for the *IBD5* variants were significantly gher in the patients with CD when compared with the ntrol population (IGR2096, 79.3% CD vs 68.7% HC = .0034]; IGR2198, 77% CD vs 67% HC [P = 11]; IGR2230, 80% CD vs 73.7% HC [P = .047]). dividuals who were homozygous for the *IBD5* risk leles at the 3 SNPs examined (IGR2096, IGR2198, d IGR2230) were more common in the CD group than the control group (Table 2). The population-attributble risk for *IBD5* homozygosity was estimated as 15% IGR2096 data were considered in calculation, 14.0% r IGR2198, and 14.3% for IGR2230.

No associations were observed between allelic freuency of *IBD5* variants and IBD overall (IGR2198, 5% IBD vs 40.8% HC [P = .077]) or UC (IGR2198, 3.7% UC vs 40.8% HC [P = .325]) (Supplementary able 2; see supplemental material online at www.mmc.ed.ac.uk/gi/supdata.shtml).

Analysis of OCTN1/2 variants and OCTN1/2-TC aplotype. Allelic frequencies differed between paents with CD and controls for the OCTN1 variant (*SLC22A4* exon 9 1672C→T) (53.6% CD vs 42.9% HC, P = .0008) and the OCTN2 variant (*SLC22A5* promoter −207G→C) (56.1% CD vs 48.4% HC, P = .0092) (Table 2). Carriage of the OCTN1/2-TC risk haplotype was present more frequently in patients with CD than in controls (80% CD vs 68.5% HC, P = .0016, OR, 1.8) (Table 2). It was clear that this difference related to homozygosity because OCTN1/2-TC homozygotes were more common in the CD group (28.4% CD vs 16.1% HC, P = .0042, OR, 1.83). No significant difference was observed between OCTN1/2-TC heterozygote rates (P = .3).

An association was observed between the OCTN1 variant and IBD when allelic frequencies were analyzed (48.9% IBD vs 42.9% HC, P = .019). This finding was not replicated in analysis of the OCTN2 variant (51.7% IBD vs 48.4% HC, P = .162), and no associations were observed between variant allelic frequencies of OCTN1 (44.7% UC vs 42.9% HC, P = .53) or OCTN2 (5% UC vs 48.4% HC, P = .38) and UC (Supplementary Table 2; see supplemental material online at www.mmc.med.ed.ac.uk/gi/supdata.shtml). No association was observed between the OCTN1/2-TC haplotype and IBD overall or with UC. There was no evidence of epistasis

able 3. Frequency of OCTN1/2 Variant Alleles and the OCTN1/2-TC Haplotype in Patients with CD Stratified by Carriage of *NOD2/CARD15* Variants

SNP/haplotype examined	% *NOD2/CARD15*-positive patients (n = 84)	% *NOD2/CARD15*-negative patients (n = 188)	*P*
CTN2 (−207 G→C) allelic frequency	55.2	56.3	.78
CTN1 (1672C→T) allelic frequency	51.9	53.4	.69
CTN1/2 TC haplotype carriage	72.3	71.1	.79
CTN1/2 TC haplotype homozygosity	22.0	24.5	.574

OTE. Analysis of evidence for epistasis between OCTN2 variant (−207G→C) and OCTN1 variant (1672C→T) and the TC haplotype of OCTN1/2 as investigated by stratifying OCTN1/2 variants by carriage of one or more of the 3 common *CARD15* variants: R702W, G908R, and 007fsinsC. No evidence of epistasis was observed.

etween the OCTN1/2 variants, the OCTN1/2-TC haplotype, and *CARD15* variants (Table 3).

The OCTN1/2 association with CD is not independent of the association with the extended *IBD5* haplotype. Previous data have suggested carriage of the OCTN1/2-TC haplotype to be an independent risk factor for CD. However, in our data set, for individuals who lacked the *IBD5* risk haplotype (homozygous with respect to the non–CD-associated alleles of IGR2096, IGR2198, and IGR2230), the OCTN1/2-TC haplotype was not associated with CD (Table 4). In the absence of variants in the marker SNP IGR2096, which is located in block 4[27] (Figure 1), 21.5% of the controls carried the OCTN1/2-TC haplotype compared with 30.4% in the CD group ($P = .22$). When the OCTN1/2-TC haplotype was analyzed in the absence of IGR2198 variants (block 5), 27.1% of the controls possessed the OCTN1/2-TC haplotype versus 17.1% of the patients with CD ($P = .13$). In the absence of variants in the marker SNP IGR2230, which is in the same block as OCTN1/2 (block 7), 3% of the controls possessed the OCTN1/2-TC haplotype versus 1.6% of the patients with CD ($P = .59$).

Phenotypic Analysis

Age at diagnosis. No association was observed between the age at diagnosis in patients with IBD overall, or UC and the IBD5 OCTN1/2 variant alleles. When allelic frequency was compared in subgroups defined by the Vienna classification for age in the patients with CD (A1, younger than 40 years; A2, older than 40 years), there was no significant difference between variant and wild-type allelic frequency for all studied SNPs.

Anatomic distribution. There was no association between the *IBD5* marker SNPs, OCTN1/2 variants, and CD when disease location was assessed by the Vienna classification (L1, terminal ileum; L2, colon; L3, ileocolonic; L4, upper gastrointestinal tract). Of note, no association was found between the presence of perianal CD

able 4. Frequency of the OCTN1/2 TC Haplotype in Individuals Not Possessing Disease Susceptibility Risk Alleles at the Markers IGR2096, IGR2198, and IGR2230

	No. of subjects not possessing IGR2096 variants	OCTN1/2-TC haplotype frequency (%)	No. of subjects not possessing IGR2198 variants	OCTN1/2-TC haplotype frequency (%)	No. of subjects not possessing IGR2230 variants	OCTN1/2-TC haplotype frequency (%)	No. of subjects not possessing IGR2096, IGR2198, and IGR2230 variants	OCTN1/2-TC haplotype frequency (%)
ntrols	79	17 (21.5)	85	23 (27.1)	66	2 (3)	63	2 (3.2)
tients with CD	69	21 (30.4)	76	13 (17.1)	62	1 (1.6)	41	0
		.22		.13		.59		.25

OTE. The number of patients with CD and controls are shown in the absence of the *IBD5* risk haplotype markers IGR2096, IGR2198, and R2230 and all 3 SNPs combined. The location of these marker SNPs within the *IBD5* locus is illustrated in Figure 1. The frequency of the CTN1/2-TC haplotype within each of these respective groups is illustrated together with *P* values. In these individuals not possessing allelic riants associated with the extended *IBD5* haplotype, the OCTN1/2-TC haplotype was not significantly associated with CD for any of the 3 arkers, providing evidence against an independent effect of the OCTN1/2 haplotype on disease susceptibility. In the data set from Peltekova al, the single marker IGR2078 (block 4) was used to define the *IBD5* risk haplotype. It is noteworthy that IGR2096, the only marker of the 3 udied in our population to show even a trend toward independent segregation from OCTN1/2, was the marker farthest from OCTN1/2 and also as closest to the single marker used by Peltekova et al.

ble 5. Phenotypic Associations of IGR2198, OCTN1/2 Variant Homozygosity, and Homozygosity for the OCTN1/2-TC Haplotype

Phenotype	Patients with CD homozygous for the variant allele/haplotype			
	IGR2198	OCTN2 (−207 G→C)	OCTN1 (1672C→T)	OCTN1/2 TC haplotype
enna classification of disease behavior at follow-up				
Inflammatory (n = 121)	14.4%	23.0%	17.4%	17.1%
Stricturing/penetrating (n = 217)	24.9%	34.8%	30.0%	29.7%
P (relative risk)	.026 (1.97)	.029 (1.78)	.011 (2.0)	.011 (2.05)
sease progression from diagnosis to follow-up				
No progression (n = 126)	14.7%	23.4%	18.2%	17.8%
Progression (n = 132)	25.9%	31.3%	26.9%	29.1%
P (relative risk)	.038 (1.72)	.18 (1.49)	.12 (1.66)	.049 (1.63)
rgery for luminal complications of CD				
No surgery (n = 137)	13.1%	21.9%	11.7%	12.3%
Surgery (n = 237)	23.2%	33.1%	26.6%	26.7%
P (odds ratio)	.037 (1.91)	.031 (1.77)	.0007 (2.7)	.0023 (2.2)

TE. An association was observed between IGR2198 variant homozygosity, OCTN2 variant homozygosity, OCTN1 variant homozygosity, and mozygosity for the OCTN1/2-TC haplotype and structuring/penetrating disease when compared with inflammatory disease at most recent llow-up. Of the 258 patients with CD who had inflammatory Vienna disease classification at diagnosis, 126 patients did not progress at llow-up and 132 patients progressed to the stricturing/penetrating group. The average duration of follow-up of these patients was 11.8 years.

d homozygosity for the OCTN1/2-TC haplotype (P = 75). When patients with CD were further stratified for *OD2/CARD15* variant carriage, no associations be- ween homozygosity for the OCTN1/2-TC haplotype d the disease location, categorized by the Vienna clas- ication, were observed; 45% of OCTN1/2-TC homozy- us patients with CD with terminal ileal disease (L1) rried no *NOD2/CARD15* variants, whereas 52% car- ed 1 or more *NOD2/CARD15* variants (P = .75). The *D5* OCTN1/2 variants were not associated with dis- se extent and severity in the UC cohort.

Disease behavior. When disease behavior at di- nosis was analyzed, no significant association between riants representing the extended *IBD5* locus, individ- l OCTN1/2 variants, and stricturing (B2, Vienna clas- ication) or penetrating (B3) behavior was observed nen compared with inflammatory, nonstricturing, non- netrating disease behavior (B1). A significant associa- on was observed between IGR2198 homozygosity and ricturing/penetrating disease when compared with in- mmatory disease at most recent follow-up (24.9% B2 d B3 vs 14.4% B1; P = .026; OR, 1.97) (Table 5). gnificant associations were also observed between stric- ring/penetrating disease when compared with inflam- atory disease for OCTN1 variant homozygosity (30% 2 and B3 vs 17.4% B1; P = .011; OR, 2.0), OCTN2 riant homozygosity (34.8% B2 and B3 vs 23.0% B1; = .029; OR, 1.78), and homozygosity for the OCTN1/ TC haplotype (29.7% B2 and B3 vs 17.1% B1; P = 11; OR, 2.05). In individuals who lacked the *IBD5* risk haplotype (homozygous with respect to the non–CD-associated allele of the IGR2198 variant), there was no association between stricturing/penetrating disease behavior at follow-up and homozygosity for the OCTN1/2-TC haplotype (11.1% inflammatory vs 14.9% stricturing/penetrating; P = .65).

Disease progression. Because the Vienna classification is a hierarchical system, a number of patients will move from inflammatory disease (B1) at diagnosis to either stricturing (B2) or penetrating (B3) disease during their follow-up. Of the 258 patients who had inflammatory disease at diagnosis, 126 patients remained in the inflammatory group at follow-up and 132 patients progressed to the stricturing group or the penetrating group. The median duration of follow-up of these patients was 11.8 years. Homozygosity rates of IGR2198 variants and the OCTN1/2-TC haplotype were associated with disease progression to B2 or B3 compared with those whose disease remained as B1 (Table 5). The association between the OCTN1/2-TC haplotype and disease progression was not observed in patients with CD who lacked the *IBD5* risk haplotype (18% progression vs 13% nonprogression; P = .56).

Requirement for surgery. As a marker of disease severity, patients who had required surgery (n = 237) for complications of luminal CD were compared with those who had had no surgery for CD (n = 137). A significant association was observed between requirement for surgery and homozygosity for variants of IGR2198,

CTN1, OCTN2, and the OCTN1/2-TC haplotype (Table 5).

Multiple logistic regression analysis was applied, using a model that considered age at diagnosis, disease behavior, smoking status, family history, *NOD2/CARD15* carrier status, and IGR2198 homozygosity or the OCTN1/2-TC haplotype, with the outcome being surgery for luminal complications of CD. Colinearity between IGR2198 and OCTN1/2-TC homozygosity was evident and allowed for in modeling. IGR2198 homozygosity ($P = .016$; odds ratio, 2.86; confidence interval, 1.21–6.76) and the OCTN1/2-TC haplotype ($P = .004$; odds ratio, 3.52; confidence interval, 1.49–8.31) were significantly associated with the need for surgery.

Discussion

The present study has provided novel data regarding the contribution of the *IBD5* locus as a determinant of disease susceptibility and phenotype in the Scottish population, which is known to be characterized by low rates of racial admixture compared with others in Europe and North America.[40] We have shown the OCTN1 and OCTN2 polymorphisms to be in tight linkage disequilibrium (D′, >0.959), with allelic variants of the 3 SNPs representing the extended *IBD5* haplotype (IGR2096, IGR2198, and IGR2230). Susceptibility to CD was associated with each of the 3 *IBD5* polymorphisms defining the extended *IBD5* region and the variant alleles of the OCTN1 and OCTN2 genes. Homozygosity for each of the 3 SNPs used to define the extended *IBD5* haplotype and for the OCTN1/2-TC haplotype was strongly associated with disease susceptibility. These data provide the first independent confirmation of the association with the OCTN1/2 variants studied by Peltekova et al in Canada.[31]

However, our data are not entirely consistent with the results reported by Peltekova et al in a potentially critical aspect. In the absence of allelic variants representing the extended risk haplotype for *IBD5*, our results showed no significant difference in the presence of the OCTN1/2-TC haplotype in patients with CD compared with controls. These findings lead to the important suggestion that the OCTN1/2-TC haplotype may not confer risk of CD independently of other closely linked determinants within the extended *IBD5* locus and that it is at present premature to conclude that the OCTN1/2 variants are causative in the pathogenesis of CD.

In comparing the studies reported by Peltekova et al with our own, the choice of markers used to define the *IBD5* locus is especially worthy of discussion. In the data set from Peltekova et al, the single marker IGR2078, located in block 4 of the haplotype map[27] (Figure 1), was used to define the *IBD5* risk haplotype and not a marker closer to the OCTN1/2 loci (block 7), such as IGR2198 (block 5) or IGR2230 (block 7), which were used in the present study. In light of this, recombination between the IGR2078 marker in block 4 and OCTN1/2 in block 7 needs to be considered as an explanation for the apparent independence reported by Peltekova et al. In our data, the only marker for which any trend toward an independence of the OCTN1/2-TC haplotype was observed concerned the IGR2096 variants, which are also located in block 4. In the absence of variants of the markers IGR2198 and IGR2230, the OCTN1/2-TC haplotype was in fact more prevalent in the control group, further supporting the hypothesis of recombination between haplotype blocks 4 and 7.[27]

Recent data presented in abstract form involving 3 large cohorts of patients with IBD (more than 1200 patients with CD in total) in Cambridge,[41] Stockholm, Sweden (Törkvist et al, personal communication, May 2005), and the United Kingdom/Germany[42] have all shown an association between the *IBD5* locus and CD. In each of these data sets, as in the present study, there was no independent association between the OCTN1/2-TC haplotype and CD in the absence of the extended *IBD5* risk haplotype.

The arguments as to the contribution of the OCTN1/2 variants implicated by Peltekova et al have been developed further recently and clearly require resolution. Studies involving 276 samples from healthy, ethnically diverse human populations have allowed the identification of several nonconservative SNPs of OCTN1/2 in evolutionary conserved sites in black and Chinese populations.[43] The investigators suggest that if these variants were shown to confer susceptibility to CD in their respective populations, this would provide strong evidence that OCTN1/2 contain the critical mutations. As yet, these studies have not been undertaken. Moreover, genetic mutations in OCTN2 cause systemic carnitine deficiency, characterized by disease of skeletal muscle, cardiac muscle, and liver but not inflammatory or intestinal disease.[44]

Genetic studies to resolve the limits of the association using markers p and q telomeric to OCTN1/2 may help to clarify this issue; however, power of resolution will become a critical issue in attempting to resolve this controversy by genetic studies alone. We have calculated that 3200 individuals with CD would need to be genotyped to prove the independence of OCTN1/2 from IGR2230, which is situated between the 2 genes in haplotype block 7.[27] Complementary approaches such as functional and expression studies in CD are clearly re-

uired. Whereas OCTN1/2 remain strong and plausible andidates within the haplotype structure on 5q31, the ase is yet to be proven beyond doubt.

Other immunoregulatory genes that have the poten- al to play a role in the pathogenesis of CD and are ocated on the *IBD5* haplotype are interleukin-3, inter- eukin-4, interleukin-5, and CSF2.[9] The interleukins ave a well-defined role in immune regulation, and CSF2 ncodes for granulocyte-macrophage colony–stimulating ctor.[9,45]

Increasingly, in CD, it is recognized that clinical henotype is genetically determined.[46] In our cohort, hen the Vienna classification for disease behavior was nalyzed at the patients' most recent follow-up assess- nent, there was a significant association between the *BD5* marker SNP IGR2198, OCTN1/2 variants, the CTN1/2-TC haplotype, and the presence of stricturing nd penetrating disease. Patients with CD who were omozygous for IGR2198 variants or homozygous for ne OCTN1/2-TC haplotype had a disease phenotype nat was more likely to progress to stricturing or pene- rating disease behavior, and multivariate analysis nowed an association among IGR2198, the OCTN1/ -TC haplotype, and the requirement for surgery, a urrogate marker for severity in CD.[47]

It is important to point out that all the genotype- henotype data contained within our study are internally onsistent, with the significant association of the *IBD5* CTN1/2 variants, stricturing and penetrating disease latest follow-up, and disease progression from inflam- natory to stricturing and penetrating disease, because nese patients with progressive complicated disease are nore likely to require surgical intervention. Although no lassification system has gained universal approval, the ienna classification is increasingly widely used and is urrently subject to reevaluation.[48] Variants within the *BD5* locus have also been associated with a more severe isease phenotype and growth failure in an independent ohort of 200 Scottish patients diagnosed with CD at 16 ears of age or younger.[49] Overall, our data show an xciting, novel genotype-phenotype association with the *BD5* locus, and it is intriguing to propose that a genetic ariant within the *IBD5* locus causes CD to become more evere (stricturing or penetrating disease) and hence re- uire increased surgical intervention.

Previous work involving the *IBD5* locus has shown an ssociation between the *IBD5* risk haplotype and peri- nal CD, but this was not replicated in our cohort.[28] henotypic data from Newman et al showed an associa- on with ileal disease,[35] but no significant association as observed in our population; 32.2% of patients with eal CD were OCTN1/2-TC homozygous and 78% carried the OCTN1/2-TC haplotype compared with 22.6% of patients with nonileal CD who were OCTN1/2 homozygotes and 75.5% of whom carried the OCTN1/2-TC haplotype ($P = .075$ and $P = .63$, respectively). Differences in definitions of ileal CD may be the reason for the differences between these 2 cohorts; ileal disease in our cohort was classified strictly according to the Vienna classification (L1). Using these criteria, 34% of patients had purely terminal ileal disease; in contrast, 70.6% of patients in the Toronto cohort were classified as having ileal disease.[35] Newman et al did not appear to include any patients with proximal small intestinal or upper gastrointestinal disease. No data were available in this cohort regarding disease behavior or disease progression over time.

NOD2/CARD15 variants have consistently been associated with a younger age of onset of disease, ileal disease, and stricturing disease.[26] In the present study, no evidence of epistasis was observed between the OCTN1/2 variants and *NOD2/CARD15* variants. Genetic heterogeneity between the Scottish population and that of Canada may be responsible for observed differences in this context, and it is pertinent to note the low frequency of *CARD15* variants in the Scottish CD population (1007fsinsC, 4.7%; G908R, 1.8%; R702W, 7.1%) and the low combined population-attributable risk these variants confer, which is estimated at 11%.[24]

In conclusion, we have determined that the *IBD5* locus is associated with disease susceptibility in our CD cohort in Scotland. This is the first independent replication of the association of the OCTN1/2 haplotype with CD outside the index population in Canada. However, a significant effect of OCTN1 and OCTN2 variants was not seen in the absence of the *IBD5* risk haplotype, and on the strength of the present data it is not possible to conclude whether or not the OCTN1/2 genes contain the disease-causing mutation or whether the association serves to only narrow the region of association within the *IBD5* locus. Finally, novel and potentially important phenotypic associations have been identified with the *IBD5* locus, which implicate this region as a determinant of disease severity in CD as well as susceptibility.

References

1. Armitage E, Drummond H, Ghosh S, Ferguson A. Incidence of juvenile-onset Crohn's disease in Scotland. Lancet 1999;353: 1496–1497.
2. Armitage EL, Aldhous MC, Anderson N, Drummond HE, Riemersma RA, Ghosh S, Satsangi J. Incidence of juvenile-onset Crohn's disease in Scotland: association with northern latitude and affluence. Gastroenterology 2004;127:1051–1057.
3. Askling J, Grahnquist L, Ekbom A, Finkel Y. Incidence of paediatric Crohn's disease in Stockholm, Sweden. Lancet 1999;354:1179.

4. Sawczenko A, Sandhu BK, Logan RF, Jenkins H, Taylor CJ, Mian S, Lynn R. Prospective survey of childhood inflammatory bowel disease in the British Isles. Lancet 2001;357:1093–1094.
5. Satsangi J, Morecroft J, Shah NB, Nimmo E. Genetics of inflammatory bowel disease: scientific and clinical implications. Best Pract Res Clin Gastroenterol 2003;17:3–18.
6. Cho JH, Nicolae DL, Gold LH, Fields CT, LaBuda MC, Rohal PM, Pickles MR, Qin L, Fu Y, Mann JS, Kirschner BS, Jabs EW, Weber J, Hanauer SB, Bayless TM, Brant SR. Identification of novel susceptibility loci for inflammatory bowel disease on chromosomes 1p, 3q, and 4q: evidence for epistasis between 1p and IBD1. Proc Natl Acad Sci U S A 1998;95:7502–7507.
7. Ma Y, Ohmen JD, Li Z, Bentley LG, McElree C, Pressman S, Targan SR, Fischel-Ghodsian N, Rotter JI, Yang H. A genome-wide search identifies potential new susceptibility loci for Crohn's disease. Inflamm Bowel Dis 1999;5:271–278.
8. Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, McLeod RS, Griffiths AM, Green T, Brettin TS, Stone V, Bull SB, Bitton A, Williams CN, Greenberg GR, Cohen Z, Lander ES, Hudson TJ, Siminovitch KA. Genomewide search in Canadian families with inflammatory bowel disease reveals two novel susceptibility loci. Am J Hum Genet 2000;66:1863–1870.
9. Rioux JD, Daly MJ, Silverberg MS, Lindblad K, Steinhart H, Cohen Z, Delmonte T, Kocher K, Miller K, Guschwan S, Kulbokas EJ, O'Leary S, Winchester E, Dewar K, Green T, Stone V, Chow C, Cohen A, Langelier D, Lapointe G, Gaudet D, Faith J, Branco N, Bull SB, McLeod RS, Griffiths AM, Bitton A, Greenberg GR, Lander ES, Siminovitch KA, Hudson TJ. Genetic variation in the 5q31 cytokine gene cluster confers susceptibility to Crohn disease. Nat Genet 2001;29:223–228.
10. Hampe J, Schreiber S, Shaw SH, Lau KF, Bridger S, MacPherson AJ, Cardon LR, Sakul H, Harris TJ, Buckler A, Hall J, Stokkers P, van Deventer SJ, Nurnberg P, Mirza MM, Lee JC, Lennard-Jones JE, Mathew CG, Curran ME. A genomewide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. Am J Hum Genet 1999;64:808–816.
11. Satsangi J, Parkes M, Louis E, Hashimoto L, Kato N, Welsh K, Terwilliger JD, Lathrop GM, Bell JI, Jewell DP. Two stage genome-wide search in inflammatory bowel disease provides evidence for susceptibility loci on chromosomes 3, 7 and 12. Nat Genet 1996;14:199–202.
12. Duerr RH, Barmada MM, Zhang L, Pfutzer R, Weeks DE. High-density genome scan in Crohn disease shows confirmed linkage to chromosome 14q11-12. Am J Hum Genet 2000;66:1857–1862.
13. Hugot JP, Laurent-Puig P, Gower-Rousseau C, Olson JM, Lee JC, Beaugerie L, Naom I, Dupas JL, Van Gossum A, Orholm M, Bonaiti-Pellie C, Weissenbach J, Mathew CG, Lennard-Jones JE, Cortot A, Colombel JF, Thomas G. Mapping of a susceptibility locus for Crohn's disease on chromosome 16. Nature 1996; 379:821–823.
14. Hugot JP, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, Binder V, Finkel Y, Cortot A, Modigliani R, Laurent-Puig P, Gower-Rousseau C, Macry J, Colombel JF, Sahbatou M, Thomas G. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 2001;411:599–603.
15. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, Britton H, Moran T, Karaliuskas R, Duerr RH, Achkar JP, Brant SR, Bayless TM, Kirschner BS, Hanauer SB, Nunez G, Cho JH. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature 2001;411:603–606.
16. Girardin SE, Boneca IG, Viala J, Chamaillard M, Labigne A, Thomas G, Philpott DJ, Sansonetti PJ. Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. J Biol Chem 2003;278:8869–8872.
17. Inohara N, Ogura Y, Fontalba A, Gutierrez O, Pons F, Crespo J, Fukase K, Inamura S, Kusumoto S, Hashimoto M, Foster SJ, Moran AP, Fernandez-Luna JL, Nunez G. Host recognition of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease. J Biol Chem 2003;278:5509–5512.
18. Watanabe T, Kitani A, Murray PJ, Strober W. NOD2 is a negative regulator of Toll-like receptor 2-mediated T helper type 1 responses. Nat Immunol 2004;5:800–808.
19. Kobayashi KS, Chamaillard M, Ogura Y, Henegariu O, Inohara N, Nunez G, Flavell RA. Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. Science 2005; 307:731–734.
20. Maeda S, Hsu LC, Liu H, Bankston LA, Iimura M, Kagnoff MF, Eckmann L, Karin M. Nod2 mutation in Crohn's disease potentiates NF-kappaB activity and IL-1beta processing. Science 2005; 307:734–738.
21. Lesage S, Zouali H, Cezard JP, Colombel JF, Belaiche J, Almer S, Tysk C, O'Morain C, Gassull M, Binder V, Finkel Y, Modigliani R, Gower-Rousseau C, Macry J, Merlin F, Chamaillard M, Jannot AS, Thomas G, Hugot JP. CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. Am J Hum Genet 2002;70:845–857.
22. Inoue N, Tamura K, Kinouchi Y, Fukuda Y, Takahashi S, Ogura Y, Inohara N, Nunez G, Kishi Y, Koike Y, Shimosegawa T, Shimoyama T, Hibi T. Lack of common NOD2 variants in Japanese patients with Crohn's disease. Gastroenterology 2002;123:86–91.
23. Leong RW, Armuzzi A, Ahmad T, Wong ML, Tse P, Jewell DP, Sung JJ. NOD2/CARD15 gene polymorphisms and Crohn's disease in the Chinese population. Aliment Pharmacol Ther 2003;17:1465–1470.
24. Arnott ID, Nimmo ER, Drummond HE, Fennell J, Smith BR, MacKinlay E, Morecroft J, Anderson N, Kelleher D, O'Sullivan M, McManus R, Satsangi J. NOD2/CARD15, TLR4 and CD14 mutations in Scottish and Irish Crohn's disease patients: evidence for genetic heterogeneity within Europe? Genes Immun 2004;5: 417–425.
25. Helio T, Halme L, Lappalainen M, Fodstad H, Paavola-Sakki P, Turunen U, Farkkila M, Krusius T, Kontula K. CARD15/NOD2 gene variants are associated with familially occurring and complicated forms of Crohn's disease. Gut 2003;52:558–562.
26. Ahmad T, Tamboli CP, Jewell D, Colombel JF. Clinical relevance of advances in genetics and pharmacogenetics of IBD. Gastroenterology 2004;126:1533–1549.
27. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. Nat Genet 2001;29:229–232.
28. Armuzzi A, Ahmad T, Ling KL, de Silva A, Cullen S, van Heel D, Orchard TR, Welsh KI, Marshall SE, Jewell DP. Genotype-phenotype analysis of the Crohn's disease susceptibility haplotype on chromosome 5q31. Gut 2003;52:1133–1139.
29. Mirza MM, Fisher SA, King K, Cuthbert AP, Hampe J, Sanderson J, Mansfield J, Donaldson P, MacPherson AJ, Forbes A, Schreiber S, Lewis CM, Mathew CG. Genetic evidence for interaction of the 5q31 cytokine locus and the CARD15 gene in Crohn disease. Am J Hum Genet 2003;72:1018–1022.
30. Giallourakis C, Stoll M, Miller K, Hampe J, Lander ES, Daly MJ, Schreiber S, Rioux JD. IBD5 is a general risk factor for inflammatory bowel disease: replication of association with Crohn disease and identification of a novel association with ulcerative colitis. Am J Hum Genet 2003;73:205–211.
31. Peltekova VD, Wintle RF, Rubin LA, Amos CI, Huang Q, Gu X, Newman B, Van Oene M, Cescon D, Greenberg G, Griffiths AM, George-Hyslop PH, Siminovitch KA. Functional variants of OCTN cation transporter genes are associated with Crohn disease. Nat Genet 2004;36:471–475.

2. Tamai I, Yabuuchi H, Nezu J, Sai Y, Oku A, Shimane M, Tsuji A. Cloning and characterization of a novel human pH-dependent organic cation transporter, OCTN1. FEBS Lett 1997;419:107–111.
3. Tokuhiro S, Yamada R, Chang X, Suzuki A, Kochi Y, Sawada T, Suzuki M, Nagasaki M, Ohtsuki M, Ono M, Furukawa H, Nagashima M, Yoshino S, Mabuchi A, Sekine A, Saito S, Takahashi A, Tsunoda T, Nakamura Y, Yamamoto K. An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. Nat Genet 2003;35:341–348.
4. Tamai I, Ohashi R, Nezu J, Yabuuchi H, Oku A, Shimane M, Sai Y, Tsuji A. Molecular and functional identification of sodium ion-dependent, high affinity human carnitine transporter OCTN2. J Biol Chem 1998;273:20378–20382.
5. Newman B, Gu X, Wintle R, Cescon D, Yazdanpanah M, Liu X, Peltekova V, Van Oene M, Amos CI, Siminovitch KA. A risk haplotype in the solute carrier family 22A4/22A5 gene cluster influences phenotypic expression of Crohn's disease. Gastroenterology 2005;128:260–269.
6. Lennard-Jones JE. Classification of inflammatory bowel disease. Scand J Gastroenterol Suppl 1989;170:2–6.
7. Gasche C, Scholmerich J, Brynskov J, D'Haens G, Hanauer SB, Irvine EJ, Jewell DP, Rachmilewitz D, Sachar DB, Sandborn WJ, Sutherland LR. A simple classification of Crohn's disease: report of the Working Party for the World Congresses of Gastroenterology, Vienna 1998. Inflamm Bowel Dis 2000;6:8–15.
8. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1988;16:1215.
9. Schlesselman JJ. Case control studies: design, conduct, analysis. In: Stolley PD, ed. Oxford: Oxford Press, 1982.
10. Capelli C, Redhead N, Abernethy JK, Gratrix F, Wilson JF, Moen T, Hervig T, Richards M, Stumpf MP, Underhill PA, Bradshaw P, Shaha A, Thomas MG, Bradman N, Goldstein DB. A Y chromosome census of the British Isles. Curr Biol 2003;13:979–984.
41. Waller S, Tremelling M, Bredin F, Padfield E, Hunter J, Middleton S, Woodward J, Parkes M. OCTN associated with Crohn's disease and ulcerative colitis but does it cause IBD? (abstr). Gut 2005;54:A17.
42. Onnie C, Fisher S, Hampe J, Mirza M, Forbes A, Mansfield J, Sanderson J, Lewis C, Schreiber S, Mathew C. Multiple SNPs at the IBD5 locus contribute to the risk of Crohn's disease (abstr). Gut 2005;54:A95.
43. Urban TJ, Giacomini KM, Risch N. Haplotype structure and ethnic-specific allele frequencies at the OCTN locus: implications for the genetics of Crohn's disease. Inflamm Bowel Dis 2005;11:78–79.
44. Lahjouji K, Mitchell GA, Qureshi IA. Carnitine transport by organic cation transporters and systemic carnitine deficiency. Mol Genet Metab 2001;73:287–297.
45. Banchereau J, Steinman RM. Dendritic cells and the control of immunity. Nature 1998;392:245–252.
46. Sachar DB. Genomics and phenomics in Crohn's disease. Gastroenterology 2002;122:1161–1162.
47. Forcione DG, Rosen MJ, Kisiel JB, Sands BE. Anti-Saccharomyces cerevisiae antibody (ASCA) positivity is associated with increased risk for early surgery in Crohn's disease. Gut 2004;53:1117–1122.
48. Arnott ID, Satsangi J. Crohn's disease or Crohn's diseases? Gut 2003;52:460–461.
49. Russell RK, Drummond H, Smith L, Anderson N, Nimmo E, Wilson DC, Gillett PM, McGrogan P, Hassan K, Weaver LT, Bisset M, Mahdi G, Satsangi J. OCTN 1/2 variants are associated with disease susceptibility and phenotype in early onset inflammatory bowel disease (IBD) (abstr). Gastroenterology 2005;128:A38.

Received November 11, 2004. Accepted September 7, 2005.

Address requests for reprints to: Jack Satsangi, DPhil, FRCP, Gastrointestinal Unit, Western General Hospital, Edinburgh, Scotland EH4 2XU. e-mail: j.satsangi@ed.ac.uk; fax: (44) 131-651-1085.

Supported by Wellcome Trust Programme grant 072789/Z/03/Z.

INFLAMMATORY BOWEL DISEASE

DLG5 variants do not influence susceptibility to inflammatory bowel disease in the Scottish population

C L Noble, E R Nimmo, H Drummond, L Smith, I D R Arnott, J Satsangi

Gut 2005;**54**:1416–1420. doi: 10.1136/gut.2005.066621

See end of article for authors' affiliations

Correspondence to: Professor J Satsangi, Gastrointestinal Unit, Western General Hospital, Edinburgh, EH4 2XU, UK; J.Satsangi@ed.ac.uk

Revised version received 4 April 2005
Accepted for publication 11 April 2005
Published online first 20 April 2005

Introduction: Recent data have suggested that specific haplotypic variants of the DLG5 gene on chromosome 10q23 may be associated with susceptibility to inflammatory bowel disease (IBD) in Germany. Haplotype D, notably characterised by the presence of a G→A substitution at nucleotide 113, was associated with susceptibility to Crohn's disease (CD) whereas an extended haplotype A conferred protection.

Aims: Association of DLG5 haplotypic variants with disease susceptibility, genotype-phenotype relationships, and epistasis with CARD15 was investigated in the Scottish population.

Patients and methods: A total of 374 CD, 305 ulcerative colitis (UC), and 294 healthy controls (HC) were studied. Genotyping for the variants rs1248696 (113A, representing haplotype D) and the single nucleotide polymorphism tag rs2289311 (representing haplotype A) were typed using the Taqman system.

Results: On analysis of the DLG5 variant 113A, there were no associations with IBD when allelic frequency (11.4% IBD *v* 13.2% HC; $p=0.30$) and carrier frequency (19.2% IBD *v* 24.6% HC; $p=0.069$) were analysed. No associations were observed between 113A variant allelic frequency ($p=0.37$), carrier frequency ($p=0.057$), and CD. In fact, 113A heterozygosity rates were lower in CD (16%) and IBD (16.9%) than in HC (23%) ($p=0.029$ and $p=0.033$, respectively). No associations between DLG5 and UC were observed. Haplotype A was not protective and there was no evidence of epistasis between DLG5 and CARD15.

Conclusions: The present data contrast strongly with previous data from Germany. DLG5 113A is not associated with disease susceptibility and haplotype A does not confer resistance. Further work is required to evaluate the significance of DLG5 in other populations from geographically diverse regions.

Crohn's disease (CD) and ulcerative colitis (UC) are common causes of gastrointestinal morbidity in the developed world, and the incidence of early onset CD has continued to rise in Northern Europe.[1–3] Increasing insight has been gained into the critical role of the disregulated immune response to bacterial flora and it is clear that genetic susceptibility to environmental agents are central to the pathogenesis of chronic intestinal inflammation.[4 5]

Genome wide scanning has identified six confirmed loci that confer susceptibility to CD and the first discovered and most consistently replicated critical mutations were found in the CARD15 (NOD2) gene on chromosome 16 (IBD1).[6 7] The physiological role of the CARD15/NOD2 protein remains under detailed examination. In vitro data suggest that CARD15 functions as an intracellular sensor of muramyl dipeptide, a highly conserved peptidoglycan motif common to many intraluminal bacteria.[8 9] These observations have now been complemented by studies of genetically engineered mouse models. Watanabe *et al* have suggested that CARD15−/− mice lose negative control of Toll-like receptor 2 (TLR2) mediated activation of nuclear factor κB, potentially offering an explanation for the Th1 phenotype characteristic of CD.[10] However, recently compiled studies do not provide support for NOD2/CARD15 interaction with the TLR2 pathway and emphasise the complexity of NOD2/CARD15 activation.[11 12]

Carriage rates of the three common NOD2/CARD15 mutations single nucleotide polymorphisms (SNPs) Gly908Arg, Arg702Trp, and the frame shift mutation Leu1007fsinsC vary between 0% and 50% in different CD cohorts, with high rates reported in Central European populations[6] and low rates in Northern Europe (Finland and Scotland).[13 14] Genetic heterogeneity between the populations of North and Central Europe is an obvious reason for the difference in CARD15 carriage rates and as more susceptibility genes are being discovered this explanation has become increasingly pertinent.[13]

A locus on chromosome 10 was first implicated in genome wide scanning of a European cohort (UK, Germany, and the Netherlands) in 1999 as conferring susceptibility to inflammatory bowel disease (IBD) (LOD score 2.30).[15] Recent data now suggest that the gene DLG5 (Drosophila discs large homologue 5) located on chromosome 10q23 may be responsible for the observed linkage and contains critical mutations that confer susceptibility to IBD.[16]

DLG5 is a member of the MAGUK (membrane associated gyanylate kinase) gene family which encode cell scaffolding proteins and are also involved in intracellular signal transduction.[17] MAGUKs interact with other proteins to create an assembly of large multiprotein complexes that bind transmembrane proteins at the cytoplasmic side to other signal transduction proteins, thus creating a platform for specific signal interactions.[18 19] The DLG5 gene spans 79 kb,

Abbreviations: IBD, inflammatory bowel disease; CD, Crohn's disease; UC, ulcerative colitis; HC, healthy controls; SNP, single nucleotide polymorphism; TLR, Toll-like receptor; DLG5, Drosophila discs large homologue 5; MAGUK, membrane associated gyanylate kinase; PSC, primary sclerosing cholangitis

contains 32 exons, and is expressed most strongly in placental tissue and less so in heart, skeletal muscle, liver, small bowel, and colon.[20]

Stoll and colleagues[16] identified two extended DLG5 haplotypes that influenced disease susceptibility in the German population. The first haplotype (named haplotype D[16]) was especially notable for the presence of a G→A substitution at nucleotide 113 that resulted in an amino change at position 30 from arginine to glutamine (R30Q). On analysis of carrier frequency, Stoll *et al* found the 113A variant to be associated with CD (25% CD *v* 17% healthy controls; $p = 0.001$) in a case control study and trends between 113A transmission and IBD ($p = 0.09$) and CD ($p = 0.065$) were observed on transmission disequilibrium testing.[16] In silico analysis suggests that the 113A (R30Q) variant may impair DLG5 scaffolding function, but as yet no expression or functional studies in IBDs have been conducted. Evidence of epistasis between the 113A variant of DLG5 and CARD15 variants was also observed in the CD cohort.[16]

The second haplotype (haplotype A[16]) was tagged by eight marker SNPs and was observed to be significantly under transmitted in the IBD group ($p = 0.006$), suggesting the haplotype may be protective.[16] No phenotypic associations were investigated and as yet no replication data have been published.

In the present study, we have assessed the contribution of the DLG5 polymorphisms rs1248696 (113A) and rs2289311 (one of the marker SNPs for the protective haplotype A)[16]) in determining genetic susceptibility to CD and UC in the Scottish population which has a high incidence of IBD. We have also investigated genotype-phenotype associations in our rigorously defined IBD population and assessed epistasis with established CARD15 mutations.

PATIENTS AND METHODS

A total of 679 patients with well characterised IBD (IBD) and 294 controls were recruited. All IBD patients attended the clinic at the Western General Hospital, Edinburgh, a tertiary referral centre for IBD in the South East of Scotland. The group comprised of 374 patients with CD and 305 with UC. The diagnosis of IBD adhered to the criteria of Lennard-Jones.[21] CD patients were classified according to the Vienna classification which involves age at diagnosis (A1, <40 years; A2, >40 years), location (L1, terminal ileum; L2 colon; L3, ileocolon; L4, upper gastrointestinal), and behaviour (B1, non-stricturing, non-penetrating; B2, stricturing; B3, penetrating).[22] UC disease severity was judged by the criteria proposed by Truelove and Witts.[23] Phenotypic data were collected by patient questionnaire, interview, and case note review, and comprised of demographics, date of onset of symptoms and diagnosis, disease location, disease behaviour, progression, extraintestinal manifestations, surgical operations, smoking history, joint symptoms, family history, and ethnicity. The study protocol was approved by Lothian Research Ethics Committee (LREC 2000/4/192).

Demographics: CD and UC

The demographics of the CD and UC patients are shown in table 1. Duration of follow up was defined as the time from diagnosis to the time of the most recent clinic review (median duration 11.8 years in the CD group and 7.5 years in the UC group). The CD group consisted of 181 males and 193 females and the UC group 171 males and 134 females, with a median age at diagnosis of 34 years.

Vienna disease classification was available for 347 (93%) CD patients at diagnosis and 374 (100%) CD patients at follow up. Full phenotypic data were available for the UC cohort.

Control subjects

A total of 294 controls (163 blood donors from the south east of Scotland and 131 healthy controls subjects) were enrolled. Allelic frequencies of DLG5 variant SNPs 113A, rs2289311, OCTN1 variant rs1050152, OCTN2 variant rs26313667, and IBD5 marker SNP IGR2198 are shown in table 2.

Genotyping

Genomic DNA was extracted from peripheral venous blood by a modified salting out technique and resuspended in 1×TE (10 mM Tris (pH 8.0), 1 M EDTA (pH 8.0)) at a final concentration of 100 ng/μl. SNPs rs1248696 (113G→A representing haplotype D) and rs2289311 (chosen because of its reliability in genotyping to represent the protective haplotype A; M Stoll, personal communication) were typed using the Taqman system. IBD patients and controls were

Table 1 Demographics and clinical features of the Crohn's disease (CD), ulcerative colitis (UC), and control groups

	Crohn's disease (n = 374)	Ulcerative colitis (n = 305)	Controls (n = 294)
Sex (M/F)	181/193	171/134	143/151
Age at diagnosis (y) (median (IQR))	27.8 (20.9–40.4)	34 (25–50)	39 (27–52)
Duration of follow up (y) (median (IQR))	11.8 (6.5–20.2)	7.5 (3.35–13.38)	
Caucasian (%)	98.7%	98.3%	
CD location at diagnosis		**UC disease extent**	
Ileal disease	125 (36%)	Proctitis 105 (34.5%)	
Colonic disease	137 (39%)	Left sided colitis 116 (37%)	
Ileal and colonic	58 (14%)	Extensive colitis 84 (27.5%)	
Upper GI disease	30 (8.5%)		
Perianal disease	75 (21.4%)		
CD location at follow up			
Ileal disease	92 (25%)		
Colonic disease	130 (35%)		
Ileal and colonic	103 (28%)		
Upper GI disease	49 (13%)		
CD behaviour at diagnosis			
Inflammatory (Vienna B1)	258 (74.8%)		
Stricturing (Vienna B2)	30 (8.4%)		
Penetrating (Vienna B3)	59 (16.8%)		
CD behaviour at follow up			
Inflammatory (Vienna B1)	142 (38%)		
Stricturing (Vienna B2)	68 (18%)		
Penetrating (Vienna B3)	164 (44%)		

IQR, interquartile range.

Table 2 Demographics and allelic frequencies in the blood transfusion samples and healthy volunteer control samples

	Blood transfusion samples	Healthy control samples
Age (y) (median (IQR)	35 (26–47)	36 (29–51)
Sex (M/F)	83/79	52/55
Allelic frequency		
113A	14.4%	12%
rs2289311	30.4%	33%
rs1050152	43.5%	42.4%
rs26313667	49%	46.1%
IGR2198	42%	40.4%

IQR, interquartile range.
Allelic frequencies of DLG5 variant single nucleotide polymorphisms (SNPs) 113A, rs2289311, OCTN1 variant rs1050152, OCTN2 variant rs26313667, and IBD5 marker SNP IGR2198 are shown to illustrate the consistency between the blood transfusion controls and the healthy volunteer controls.

typed for polymorphisms of the CARD15 gene (R702W, G908R, and 1007fsinsC) using previously described methods.[13] All genotyping except R702W was carried out using the Taqman system. R702W genotyping was performed by restriction fragment length polymorphism polymerase chain reaction. Restriction digestion was preformed using 1 u Msp*I* at 37°C overnight and polymerase chain reaction fragments run on 4% NewSieve 3:1 agarose gels. These were stained with ethidium bromide and viewed under ultraviolet light. An image was recorded digitally.

Data analysis

The two SNPs rs1248696 and rs2289311 were analysed for association with IBD overall, CD, UC, and disease phenotype. Allelic frequency, carrier frequency, heterozygosity, and homozygosity rates were studied. Each allele was shown to be in Hardy-Weinberg equilibrium. Genotype-phenotype associations were analysed by χ^2 analysis using the Minitab statistical software package version 13/02 (Minitab Ltd, Coventry, UK). To identify significant independent variables associated with genotype, univariate and multivariate analysis was carried out. Evidence for DLG5 epistasis with CARD15 was investigated by stratifying DLG5 variants by carriage of one or more of the three common CARD15 variants—R702W, G908R, and 1007fsinsC. Allelic frequencies of the DLG5 variants were compared between the subgroups of patients with and without CARD15 variants by χ^2 analysis. The null hypothesis was that the frequency of DLG5 variants did not differ between these subgroups. Phenotypic associations of DLG5 variants were also stratified for the presence and absence of CARD15 variants.

RESULTS

Disease susceptibility: haplotype D (113A)

On analysis of the DLG5 variant 113A there were no associations with IBD when allelic frequency (11.4% IBD *v* 13.2% healthy controls (HC); p = 0.30), carrier frequency (19.2% IBD *v* 24.6% HC; p = 0.069), and homozygosity rates (2.3% IBD *v* 1.5% HC; p = 0.48) were analysed (table 3). A negative association was observed between heterozygous rates of 113A and IBD (16.9% IBD *v* 23% HC; p = 0.033). Furthermore, a negative correlation was observed between heterozygous 113A variants and CD (16% CD *v* 23% HC; p = 0.029). No associations were observed between 113A variant allelic frequency (11.4% CD *v* 13.2% HC; p = 0.37), carrier frequency (18.3% CD *v* 24.6% HC; p = 0.057), homozygous rates (p = 0.36), and CD. No associations between 113A and UC were observed—allelic frequency (12.8% UC *v* 13.2% HC; p = 0.34), carrier frequency (20.3% UC *v* 24.6% HC; p = 0.23), heterozygous rates (18% UC *v* 23% HC; p = 0.45), and homozygous rates (2.3% UC *v* 1.5% CD; p = 0.5).

DLG5 haplotype A

Haplotype A allelic frequencies, represented by rs2289311 variants, did not differ between HC (31.5%) and IBD (35%; p = 0.17), CD (36.9%; p = 0.078), or UC (33.4%; p = 0.51) patients. No significant differences were observed between carriage rates of rs2289311 variants—HC (52%), IBD (57.2%; p = 0.18), CD (60.9%; p = 0.052), and UC (54%; p = 0.65). The frequency of patients who were heterozygotes or homozygotes for rs2289311 polymorphisms did not differ between IBD and HC groups (heterozygote HC 41.1% *v* IBD 44.8% (p = 0.43), CD 48% (p = 0.13), and UC 42.1% (p = 0.83); homozygote HC 10.9% *v* IBD 12.6% (p = 0.49), CD 12.6% (p = 0.61), and UC 12.9% (p = 0.50)).

Phenotypic analysis

On univariate analysis of CD patients, no association was observed between DLG5 113A variants and the Vienna classification for age of diagnosis, location of disease, or disease behaviour. Location of disease and disease behaviour in CD patients was analysed at the time of diagnosis and at the most recent follow up and there was no association between DLG5 113A variants and disease progression. No association was observed between DLG5 113A variants and age at diagnosis in the IBD and UC groups and there was no association between DLG5 113A variants, disease extent, and severity in UC patients.

DLG5 113A variants displayed a trend towards being less common in IBD patients with joint problems (large joint arthralgias related to disease activity, small joint arthralgias unrelated to disease activity, ankylosing spondylitis, and sacroilitis) (n = 127) compared with those who had no joint problems when allelic frequency was analysed (7.5% *v* 11.8%;

Table 3 DGL5 113A variant allele frequency, carrier frequency, heterozygote frequency, and homozygote frequency in the inflammatory bowel disease, Crohn's disease, ulcerative colitis, and control populations

	Controls	IBD (p value)	Crohn's disease (p value)	Ulcerative colitis (p value)
Allelic frequency	13.2%	11.4% p=0.30	11.4% p=0.37	11.4% p=0.34
Carrier frequency	63/256 (24.6%)	125/652 (19.2%) p=0.069	65/356 (18.3%) p=0.057	60/296 (20.3%) p=0.228
Heterozygosity rates	59/256 (23%)	110/650 (16.7%) p=0.033	57/356 (16%) p=0.029	53/294 (18%) p=0.45
Homozygosity rates	4/256 (1.5%)	15/650 (2.3%) p=0.48	8/356 (2.2%) p=0.36	7/294 (2.4%) p=0.50

The p values shown are calculated between the control group and each respective disease group.

$p = 0.053$). When allelic frequency of 113A variants was analysed in UC patients with primary sclerosing cholangitis, a trend towards these patients having fewer 113A variants was observed (0% ($n = 7$) *v* 11.4%; $p = 0.17$). On analysis of the haplotype A, no genotype-phenotype associations were observed in the CD and UC patient groups. Multiple logistic regression analysis did not identify any variables that were independently associated with haplotype D (113A) or haplotype A.

There was no evidence of epistasis between DLG5 113A variants and carriage of the three common CARD15 variants Gly908Arg, Arg702Trp, and Leu1007fsinsC—CARD15 carriage positive DLG5 113A allelic frequency 9.7% ($n = 108$) versus CARD15 carriage negative DLG5 113A allelic frequency 11.6% ($n = 584$) ($p = 0.43$). When CD patients were stratified for CARD15 variant carriage, no significant genotype-phenotype relationships were found with DLG5 113A.

DISCUSSION

The present study has demonstrated that in the Scottish population, traditionally characterised by low rates of admixture, the DLG5 variant 113A, representing haplotype D, is not a critical determinant of susceptibility in either CD or UC. In fact, heterozygous rates of 113A were significantly higher in the healthy control population when compared with the IBD cohort and the CD cohort. Haplotype A represented by the SNP rs2289311 was not protective in our CD or UC population.

These data differ markedly from those of Stoll *et al* who showed that in a German population, DLG5 113A variants were overtransmitted to individuals with IBD, and in a case control study there were significantly higher rates of 113A carriage in the IBD group compared with the control group (25% *v* 17%; $p = 0.001$).[16] Considerations, including sample size and phenotypic differences, between the present study and that of Stoll *et al*, may be responsible for the observed discrepancy in results but a more plausible explanation would be genetic heterogeneity between the populations of Germany and Scotland. This has been clearly illustrated by data now available with respect to the three common polymorphisms of the NOD2/CARD15 gene (G908R, R702W, and 100fsinsC) which are significantly more common in the Central European CD population[6 24] than in the Northern European CD population.[13 14 25] Furthermore, CARD15 polymorphisms are absent in Japanese and Chinese CD populations.[26–28] The different incidences of the R702W polymorphism has also been shown in healthy volunteers in Europe, Africa, and Asia.[29] In the eight independent groups worldwide who have performed genome wide scans in IBD patients, chromosome 10 has not met the stringent criteria for significant linkage.[30 31] This would suggest that if indeed DLG5 plays a role in the pathogenesis of IBD, its contribution may be limited to specific populations.

Further data illustrating genetic heterogeneity in European IBD patients have been observed on analysis of the Asp299Gly mutation of the TLR4 gene.[32] TLR4 is a member of the Toll-like receptor family which are involved in recognition of pathogen associated molecular patterns by the immune system and TLR4 functions as an extracellular pattern recognition receptor for lipopolysaccharide which is common to many intraluminal bacteria.[33] The Asp229Gly variant has been shown to confer susceptibility to CD and UC in the Belgian population[34] but no association was observed between Asp229Gly variants and IBD in German and Scottish cohorts.[13 35]

Univariate and multivariate phenotypic analysis showed no associations with DLG5 113A variants (haplotype D) and the Vienna classification of CD or with age of onset of disease or disease severity in UC. Trends were observed towards a lower DLG5 113A variant frequency in CD patients with arthropathy and UC patients with primary sclerosing cholangitis (PSC) on univariate analysis. Both of these extraintestinal complications have been shown to have a molecular genetic basis. Susceptibility to axial arthropathy has been strongly associated with HLA-B*27 in patients with and without IBD,[36] and peripheral arthropathies in patients with IBD have been associated with HLA-DRB1*0103, HLA-B*35, HLA-B*27, and HLA-B*44.[37] In PSC, strong disease associations with extended HLA haplotypes have been observed,[38] and a functional variant of stromolysin (matrix metalloproteinase 3) has also been associated with susceptibility to PSC and with disease progression.[39] In the current investigation, the relatively small numbers of IBD patients studied with these specific extraintestinal manifestations mean that these results regarding any DLG5 effect in these subgroups should be regarded as exploratory observations. Replication studies in other cohorts may help shed some light on this question.

No evidence of epistasis between DLG5 113A variants and carriage of the three common CARD15 variants Gly908Arg, Arg702Trp, and Leu1007fsinsC was observed in patients with CD. Again, these data contrast with Stoll *et al* who found significantly greater transmission of DLG5 113A in patients with CD carrying one of the risk associated alleles of CARD15.[16] CARD15 variants have consistently been associated with a younger age of onset of disease, ileal disease, and stricturing disease.[40] A possible explanation for the absence of epistasis between DLG5 and CARD15 could be the low incidence of CARD15 variants in the Scottish CD population (1007fsinsC = 4.7%, G908R = 1.8%, and R702W = 7.1%) and the combined population attributable risk these variants confer (11%).[13] Furthermore, Moore has illustrated the problems in using statistical epistasis to interpret genetic and biological phenomena.[41]

Although the present data suggest the these DLG5 variants are not important determinants in the Scottish IBD population, our own recent studies suggest that NOD2/CARD15, MDR1, and IBD5 variants are involved in disease susceptibility and behaviour.[13 42] The identity of other genetic determinants in the Northern European IBD population remains under detailed investigation.

In conclusion, in our North European study population, we were unable to replicate Stoll's data that the DLG5 variant 113A confers susceptibility to IBD. Haplotype A represented by the SNP rs2289311 did not confer protection in our population. Further genetic studies of DLG5 polymorphisms in IBD populations are required to elucidate whether these variants play a role in the pathogenesis of IBD. These studies must be complemented by data regarding the expression and function of DLG5 in the gastrointestinal tract.

ACKNOWLEDGEMENT

This project was supported by a Wellcome Trust Programme Grant: 072789/Z/03/Z.

Authors' affiliations

C L Noble, E R Nimmo, H Drummond, L Smith, I D R Arnott, J Satsangi, Gastrointestinal Unit, University of Edinburgh, Western General Hospital, Edinburgh, UK

Conflict of interest: None declared.

REFERENCES

1 **Armitage E**, Drummond H, Ghosh S, *et al*. Incidence of juvenile-onset Crohn's disease in Scotland. *Lancet* 1999;**353**:1496–7.
2 **Armitage EL**, Aldhous MC, Anderson N, *et al*. Incidence of juvenile-onset Crohn's disease in Scotland: association with northern latitude and affluence. *Gastroenterology* 2004;**127**:1051–7.
3 **Askling J**, Grahnquist L, Ekbom A, *et al*. Incidence of paediatric Crohn's disease in Stockholm, Sweden. *Lancet* 1999;**354**:1179.

4 **Bonen DK**, Cho JH. The genetics of inflammatory bowel disease. *Gastroenterology* 2003;**124**:521–36.
5 **Satsangi J**, Morecroft J, Shah NB, *et al*. Genetics of inflammatory bowel disease: scientific and clinical implications. *Best Pract Res Clin Gastroenterol* 2003;**17**:3–18.
6 **Hugot JP**, Chamaillard M, Zouali H, *et al*. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001;**411**:599–603.
7 **Ogura Y**, Bonen DK, Inohara N, *et al*. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 2001;**411**:603–6.
8 **Girardin SE**, Boneca IG, Viala J, *et al*. Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. *J Biol Chem* 2003;**278**:8869–72.
9 **Inohara N**, Ogura Y, Fontalba A, *et al*. Host recognition of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease. *J Biol Chem* 2003;**278**:5509–12.
10 **Watanabe T**, Kitani A, Murray PJ, *et al*. NOD2 is a negative regulator of Toll-like receptor 2-mediated T helper type 1 responses. *Nat Immunol* 2004;**5**:800–8.
11 **Kobayashi KS**, Chamaillard M, Ogura Y, *et al*. Nod2-dependent regulation of innate and adaptive immunity in the intestinal tract. *Science* 2005;**307**:731–4.
12 **Maeda S**, Hsu LC, Liu H, *et al*. Nod2 mutation in Crohn's disease potentiates NF-kappaB activity and IL-1beta processing. *Science* 2005;**307**:734–8.
13 **Arnott ID**, Nimmo ER, Drummond HE, *et al*. NOD2/CARD15, TLR4 and CD14 mutations in Scottish and Irish Crohn's disease patients: evidence for genetic heterogeneity within Europe? *Genes Immun* 2004;**5**:417–25.
14 **Helio T**, Halme L, Lappalainen M, *et al*. CARD15/NOD2 gene variants are associated with familially occurring and complicated forms of Crohn's disease. *Gut* 2003;**52**:558–62.
15 **Hampe J**, Schreiber S, Shaw SH, *et al*. A genomewide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. *Am J Hum Genet* 1999;**64**:808–16.
16 **Stoll M**, Corneliussen B, Costello CM, *et al*. Genetic variation in DLG5 is associated with inflammatory bowel disease. *Nat Genet* 2004;**36**:476–80.
17 **Gonzalez-Mariscal L**, Betanzos A, Avila-Flores A. MAGUK proteins: structure and role in the tight junction. *Semin Cell Dev Biol* 2000;**11**:315–24.
18 **Dimitratos SD**, Woods DF, Stathakis DG, *et al*. Signaling pathways are focused at specialized regions of the plasma membrane by scaffolding proteins of the MAGUK family. *Bioessays* 1999;**21**:912–21.
19 **Stehle T**, Schulz GE. Refined structure of the complex between guanylate kinase and its substrate GMP at 2.0 A resolution. *J Mol Biol* 1992;**224**:1127–41.
20 **Shah G**, Brugada R, Gonzalez O, *et al*. The cloning, genomic organization and tissue expression profile of the human DLG5 gene. *BMC Genomics* 2002;**3**:6.
21 **Lennard-Jones JE**. Classification of inflammatory bowel disease. *Scand J Gastroenterol Suppl* 1989;**170**:2–6.
22 **Gasche C**, Scholmerich J, Brynskov J, *et al*. A simple classification of Crohn's disease: report of the Working Party for the World Congresses of Gastroenterology, Vienna 1998. *Inflamm Bowel Dis* 2000;**6**:8–15.
23 **Truelove SC**, Witts LJ. Cortisone in ulcerative colitis; final report on a therapeutic trial. *BMJ* 1955;**4947**:1041–8.
24 **Lesage S**, Zouali H, Cezard JP, *et al*. CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. *Am J Hum Genet* 2002;**70**:845–57.
25 **Bairead E**, Harmon DL, Curtis AM, *et al*. Association of NOD2 with Crohn's disease in a homogenous Irish population. *Eur J Hum Genet* 2003;**11**:237–44.
26 **Inoue N**, Tamura K, Kinouchi Y, *et al*. Lack of common NOD2 variants in Japanese patients with Crohn's disease. *Gastroenterology* 2002;**123**:86–91.
27 **Leong RW**, Armuzzi A, Ahmad T, *et al*. NOD2/CARD15 gene polymorphisms and Crohn's disease in the Chinese population. *Aliment Pharmacol Ther* 2003;**17**:1465–70.
28 **Sugimura M**, Kinouchi Y, Takahashi S, *et al*. CARD15/NOD2 mutational analysis in Japanese patients with Crohn's disease. *Clin Genet* 2003;**63**:160–2.
29 **Marsh S**, McLeod HL. Crohn's disease: ethnic variation in CARD15 genotypes. *Gut* 2003;**52**:770.
30 **Lander E**, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 1995;**11**:241–7.
31 **Zheng CQ**, Hu GZ, Zeng ZS, *et al*. Progress in searching for susceptibility gene for inflammatory bowel disease by positional cloning. *World J Gastroenterol* 2003;**9**:1646–56.
32 **Arnott ID**, Ho GT, Nimmo ER, *et al*. Toll-like receptor 4 gene in IBD: further evidence for genetic heterogeneity in Europe. *Gut* 2005;**54**:308–9.
33 **Barton GM**, Medzhitov R. Toll-like receptor signaling pathways. *Science* 2003;**300**:1524–5.
34 **Franchimont D**, Vermeire S, El Housni H, *et al*. Deficient host-bacteria interactions in inflammatory bowel disease? The toll-like receptor (TLR)-4 Asp299gly polymorphism is associated with Crohn's disease and ulcerative colitis. *Gut* 2004;**53**:987–92.
35 **Torok HP**, Glas J, Tonenchi L, *et al*. Polymorphisms of the lipopolysaccharide-signaling complex in inflammatory bowel disease: association of a mutation in the Toll-like receptor 4 gene with ulcerative colitis. *Clin Immunol* 2004;**112**:85–91.
36 **Brewerton DA**, Caffrey M, Nicholls A, *et al*. HL-A 27 and arthropathies associated with ulcerative colitis and psoriasis. *Lancet* 1974;**1**:956–8.
37 **Orchard TR**, Thiyagaraja S, Welsh KI, *et al*. Clinical phenotype is related to HLA genotype in the peripheral arthropathies of inflammatory bowel disease. *Gastroenterology* 2000;**118**:274–8.
38 **Spurkland A**, Saarinen S, Boberg KM, *et al*. HLA class II haplotypes in primary sclerosing cholangitis patients from five European populations. *Tissue Antigens* 1999;**53**:459–69.
39 **Satsangi J**, Chapman RW, Haldar N, *et al*. A functional polymorphism of the stromelysin gene (MMP-3) influences susceptibility to primary sclerosing cholangitis. *Gastroenterology* 2001;**121**:124–30.
40 **Ahmad T**, Armuzzi A, Bunce M, *et al*. The molecular classification of the clinical manifestations of Crohn's disease. *Gastroenterology* 2002;**122**:854–66.
41 **Moore JH**. A global view of epistasis. *Nat Genet* 2005;**37**:13–14.
42 **Ho GT**, Nimmo ER, Tenesa A, *et al*. Allelic variations of the multidrug resistance gene determine susceptibility and disease behavior in ulcerative colitis. *Gastroenterology* 2005;**128**:288–96.

Hopkins Endoscopy Prize 2006

Applications are invited by the Endoscopy Committee of the British Society of Gastroenterology who will recommend to Council the recipient of the 2006 Award. Applications (10 copies) should include:

- A manuscript (two A4 pages **only**) describing work conducted
- A bibliography of relevant personal publications
- An outline of the proposed content of the lecture, including title
- A written statement confirming that all or a substantial part of the work has been personally conducted in the UK or Eire.

An applicant need not be a member of the Society. The recipient will be required to deliver a 20 minute lecture at the annual meeting of the Society in Birmingham in March 2006. Applications (10 copies) should be made to the Endoscopy Section Secretary, British Society of Gastroenterology, 3 St Andrews Place, London NW1 4LB by 1 December 2005.